University of Plymouth

PEARL **https://pearl.plymouth.ac.uk**

04 University of Plymouth Research Theses 01 Research Theses Main Collection

2008

Computational and Psycho-Physiological Investigations of Musical Emotions

Coutinho, Eduardo

http://hdl.handle.net/10026.1/1742

http://dx.doi.org/10.24382/1502
University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

Computational and Psycho-Physiological Investigations of Musical Emotions

Eduardo Coutinho

Adaptive Behaviour & Cognition Research Group

University of Plymouth

A thesis submitted to the University of Plymouth in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

October 2008

University of Plymouth
Library
9008478407
Bmark
THESIS 152.4 COU

Computational and Psycho-Physiological Investigations of Musical Emotions

Eduardo Coutinho

The ability of music to stir human emotions is a well known fact (Gabrielsson & Lindström, 2001). However, the manner in which music contributes to those experiences remains obscured. One of the main reasons is the large number of syndromes that characterise emotional experiences. Another is their subjective nature: musical emotions can be affected by memories, individual preferences and attitudes, among other factors (Scherer & Zentner, 2001). But can the same music induce similar affective experiences in all listeners, somehow independently of acculturation or personal bias? A considerable corpus of literature has consistently reported that listeners agree rather strongly about what type of emotion is expressed in a particular piece or even in particular moments or sections (Juslin & Sloboda, 2001). Those studies suggest that music features encode important characteristics of affective experiences, by suggesting the influence of various structural factors of music on emotional expression. Unfortunately, the nature of these relationships is complex, and it is common to find rather vague and contradictory descriptions.

This thesis presents a novel methodology to analyse the dynamics of emotional responses to music. It consists of a computational investigation, based on spatiotemporal neural networks sensitive to structural aspects of music, which "mimic" human affective responses to music and permit to predict new ones. The dynamics of emotional responses to music are investigated as computational representations of perceptual processes (psychoacoustic features) and self-perception of physiological activation (peripheral feedback). Modelling and experimental results provide evidence suggesting that spatiotemporal patterns of sound resonate with affective features underlying judgements of subjective feelings. A significant part of the listener's affective response is predicted from the a set of six psychoacoustic features of sound - tempo, loudness, multiplicity (texture), power spectrum centroid (mean pitch), sharpness (timbre) and mean STFT flux (pitch variation) - and one physiological variable - heart rate. This work contributes to new evidence and insights to the study of musical emotions, with particular relevance to the music perception and emotion research communities.

Contents

Abstract. v

List of Tables. xiii

List of Illustrations. xx

Acknowledgements. xxii

Author's declaration. xxv

1 Introduction 1

1.1 Scope of research 4

1.2 Organisation of this thesis 5

2 The nature of emotion in music 7

2.1 A substrate for emotions 8

2.2 The modalities of Emotion 10

2.2.1 Subjective feeling 11

2.2.2 Behavioural changes 12

2.2.3 Physiological arousal 13

2.3 The nature of Emotion in Music 16

2.3.1 "Emotivist" vs. "Cognitivist" 16

2.3.2 Music elements and the expression of emotion 22

2.4 Measurements of musical emotions 24

2.4.1 Subjective feeling measurements 25

2.4.2 Physiological measurements 30

3 Computational modelling in music and emotions 35

3.1 The use of computational models in music 36

3.1.1 Creativity and improvisation 37

3.1.2 Computational Musicology 38

3.1.3 Expression in music performance 39
3.1.4 Musical instruments and emotion 40
3.1.5 Classification of music selections 41
3.1.6 Controlling music emotionality 43
3.1.7 Emotional responses and electrical activity in the brain . . . 44
3.1.8 Time series analysis of music and emotions 45
3.2 Connectionism and Artificial Neural Networks 47
3.2.1 Processing units 49
3.2.2 Topologies 51
3.2.3 Training and learning in neural networks 51
3.3 Spatiotemporal connectionist models 52
3.3.1 Motivation for the use of Elman Neural Networks 55

4 Modelling subjective feelings **57**
4.1 Simulations methodology 57
4.1.1 Music pieces 58
4.1.2 Psychoacoustic encoding (model input data) 58
4.1.3 Experimental data on subjective feelings (model output data) 61
4.1.4 Simulation procedure 61
4.2 Reduction of the psychoacoustic (input) dimensions 65
4.2.1 Testing individual sound features 65
4.2.2 Selected sound features: model inputs 67
4.3 Analysis of model performance 68
4.3.1 Model internal dynamics: discriminant functions 73
4.3.2 Lesioning tests: long-term memory analysis 76
4.3.3 Input/output transformation: model production rules 80
4.3.4 Summary 84
4.4 Discussion and Conclusions 86

5 Psychophysiological investigation **89**
5.1 Experimental Study 91
5.2 Method 92
5.2.1 Participants 92
5.2.2 Equipment 92
5.2.3 Stimuli 93
5.2.4 Psychoacoustic encoding 97
5.2.5 Procedure 99
5.2.6 Experiment design 100

5.3 Results . . . 101
5.3.1 Data processing methods . . . 101
5.3.2 Analysis of whole music pieces . . . 112
5.3.3 Analysis of music segments . . . 114
5.3.4 Analysis of continuous measurements . . . 119
5.3.5 Self-report discrimination from sound features and physiological activity . . . 131
5.4 Discussion . . . 134

6 Modelling subjective feelings and physiological arousal 137
6.1 Simulation Methodology . . . 139
6.1.1 Music pieces . . . 139
6.1.2 Psychoacoustic encoding (model input data) . . . 139
6.1.3 Simulation procedure . . . 141
6.2 Model 1: Modelling musical emotions with sound features . . . 143
6.2.1 Hidden layer size . . . 144
6.2.2 Testing new sound features (model inputs) . . . 145
6.2.3 Model performance . . . 146
6.2.4 Comparison with the previous model . . . 147
6.3 Model 2: modelling musical emotions with sound features and physiological cues . . . 148
6.3.1 Hidden layer size . . . 149
6.3.2 New input dimension: physiological variables . . . 151
6.3.3 Model 2 performance . . . 151
6.4 Models 1 and 2 comparison . . . 152
6.5 Heart Rate and subjective feelings . . . 153
6.5.1 Lesioning tests: long-term memory analysis . . . 157
6.5.2 Model internal dynamics and input/output transformation . 160
6.6 Discussion . . . 162

7 Conclusions and future research 167
7.1 Summary and conclusions . . . 167
7.2 Contribution to knowledge . . . 172
7.3 Future research . . . 174

A Experiment: call for participants 177

B Questionnaire 179

C Participants Information Sheet 183

D IAPS pictures 187

E Randomised order of pieces presentation 189

F Sound features visualisation 191

Glossary. 201

List of references. 201

Bound copies of published papers. 221

List of Tables

4.1 Pieces used in Korhonen's experiment and their aliases for reference in this paper. The pieces were taken from Naxos's "Discover the Classsics" CD 8.550035-36 58

4.2 Psychoacoustic variables considered for this study. The aliases indicated will be used in this article to refer to the variables they refer to in this table. 59

4.3 rms error for each input data set using a model with 5 hidden units. The values shown were averaged across 3 replications of each simulation test case. For comparison purposes, the mean rms error of both outputs for a network with random weights was was established as reference value ($rms_{random} = 0.107$). 67

4.4 Comparison between the model outputs and experimental data: root mean square (rms) error and linear correlation coefficient (r) ($^{*}p < 0.0001$, $^{**}p < 0.001$, $^{***}p < 0.02$) 69

4.5 Factor Structure Matrix: correlations between discriminant variables and each hidden unit. 76

4.6 Canonical Correlation Analysis (CCA): the canonical correlations (the canonical correlations are interpreted in the same way as the Pearson's linear correlation coefficient) quantify the strength of relationships between the extracted canonical variates, and so the significance of the relationship. To assess the relationship between the original variables (input, hidden and output units activity) and the canonical variables, the canonical loadings (the correlations between the canonical variates and the variables in each set) are also included. 82

5.1 Pieces used in the experiment. 94

5.2 Psychoacoustic variables considered for this study. For convenience the input variables will be referred to with the aliases indicated in the table throughout this chapter. 98

5.3 Mean and standard deviation values for arousal, valence, heart rate and galvanic skin response across all participants for each pieces. 114
5.4 Number of segments and correspondent sound file times for each piece. 116
5.5 Correlations between sound features (L, T, P, C, mP, Tx, R, S and TW) and physiological activity (rHR - relative value of the Heart Rate level; rSCR - relative value of the Skin Conductance Response). * $p < 0.05$ 118
5.6 Correlations between sound features (L, T, P, C, mP, Tx, R, S and TW) and physiological activity (Arousal and Valence). ** $p < 0.01$, * $p < 0.05$ 119
5.7 Number of peaks in each variable per music. 121
5.8 Correlations between number of peaks in each variable. * $p < 0.001$ 121
5.9 Number of strong changes in self report within 1-5s of physiological events. 131

6.1 Aliases for the pieces used in the experiment. See Table 5.1 for further details 139
6.2 Psychoacoustic variables considered for this study. 140
6.3 Model 1: average rms errors over 15 trials for all 7 simulations with different number of hidden nodes. The mean values across the training and the test data sets are also shown for each model. . . . 145
6.4 Model 1: simulations with new sound features; the rms errors correspond to the average of the fifteen trials. 145
6.5 MODEL 1: rms errors and r coefficient, per variable, for each music piece for best trial of simulation $M_1 I_0$. * $p < 0.0001$ 146
6.6 Comparison between average rms errors and r correlation coefficient for Model 0 (Chapter 4) and Model 1 (this chapter). . . . 147
6.7 Model 2: rms errors for 7 simulations with different number of hidden nodes. The mean values across the training and the test data sets are also shown for each model. 149
6.8 Model 2: simulations with different combinations of physiological variable inputs (average rms errors over all trials). 151
6.9 Model 2: rms errors and r coefficient, per variable, for each music piece for best trial of simulation $M_2 I_0$. * $p < 0.0001$ 152

6.10 Comparison between Models 1 and 2: rms errors and linear correlation coefficient (r). These values are the mean values across all pieces for each of the models. The details for each are shown in Tables 6.5 (Model 1) and 6.9 (Model 2). 153

6.11 Canonical Correlation Analysis (CCA): the canonical correlations (the canonical correlations are interpreted in the same way as the Pearson's linear correlation coefficient) quantify the strength of relationships between the extracted canonical variates, and so the significance of the relationship. To assess the relationship between the original variables (inputs and hidden units activity) and the canonical variables, the canonical loadings (the correlations between the canonical variates and the variables in each set) are also included. 161

D.1 Pictures used in the experiment to test participants understanding of of the 2-dimensional affective space of Arousal and Valence. The pictures in the table were obtained from the International Affective Picture System (IAPS) database. 187

E.1 Number of times each music was played in which order during the experiment. 189

List of Illustrations

2.1 Taxonomy of self report measures of emotional response to music. Adapted from Schubert (1999a, p. 34) 25

3.1 Typical activation functions used in the units (artificial neurons) of artificial neural networks. 50

4.1 Areas covered by the pieces grouped by training (a) and test (b) sets. In both plots, the colours of each cloud of points correspond to all the arousal/valence pairs from each piece (see legend). Each point corresponds to the specific location on the 2DES of the arousal and valence values on a second by second basis. 62

4.2 Schematic of an Elman neural network as used in simulations. It consists on a variation on the multilayer perceptron, with an extra memory layer, which provides the network with feedback connections. The inputs to the network are the sound (psychoacoustic) features, and the outputs the affective dimensions under investigation (arousal and valence). 64

4.3 Neural network architecture and units identification for Model 0 (model used in simulation experiments): Input units - sound features (T, Tx, L, P, S and Pv); Hidden units - H_1 to H_5; Memory (context) units - M_1 to M_5; Output units - arousal (A) and valence (V). 68

4.4 Training pieces - Arousal and Valence model outputs compared to experimental data for the training data set: a) Piece 1 (Rodrigo, *Concierto de Aranjuez*), b) Piece 2 (Copland, *Fanfare for the Common Man*) and c) Piece 5 (Strauss, *Pizzicato Polka*). 70

4.5 Test pieces - Arousal and Valence model outputs compared to experimental data for the test data set: a) Piece 3 (Beethoven, *Moonlight Sonata*), b) Piece 4 (Grieg, *Peer Gynt Suite No 1*) and c) Piece 6 (Liszt, *Piano Concerto no.1*). 71

4.6 Linear discriminant analysis of hidden layer activations (canonical discriminant functions plot). The labels Q_1, Q_2, Q_3 and Q_4 indicate the quadrant in the 2DES that each coloured cluster represents. The plot shows that the spatiotemporal structures detected by the model are organised as a 2-dimensional space, and differentiated in terms of arousal and valence. 75

4.7 Neural network weight connection matrices: memory to hidden layers (top), input to hidden layers (middle) and hidden to output layers (bottom). The weights are represented as rectangles of variable size and colour: the size is proportional to the weight value and the colour represents the signal of the weight (red for negative and green for positive). 77

4.8 Model weight matrices analysis: each learned weight in the model was removed (value set to 0.0, one at a time), and the model performance was then measured using the rms error. Each cell in the above matrices corresponds to the removal of one connection linking two processing units, and the values indicated in each cell correspond to the rms error. For easier reading, the rms errors are represented using a colour code: black for those weights that had small or no effect on the model performance ($rms < 0.09$); for higher errors grey ($0.09 < rms < 0.30$) and white were used ($rms >= 0.30$). 79

4.9 Qualitative representation of individual relationships between music variables and emotion appraisals: summary of observations from model analysis. The direction of the arrows indicates an increase in the variable indicated (the arrow sizes and angles formed with both axis are merely qualitative, and cannot be interpreted in mathematical terms) 85

5.1 Experiment questionnaires: The participants' answers to each question are depicted by the smallest observation, lower quartile, median, upper quartile, and largest observations. 93

5.2 Experimental framework: participants listened to the music using a pair of closed headphones. In front of them, a computer screen shows the EMuJoy interface for the self report of Arousal and Valence. Leads were attached to measure Heart Rate and Skin Conductance Response. 99

5.3 Albinoni, *Adagio* (Piece 1): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants. 103
5.4 Grieg, *Peer Gynt Suite No. 1* (Piece 2): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants. . . . 104
5.5 Bach, *Prelude and Fugue No. 15* (Piece 3) : mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants. 105
5.6 Beethoven, *Romance No. 2* (Piece 4): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants. . . . 106
5.7 Chopin, *Nocturne No. 2* (Piece 5): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants. 107
5.8 Mozart, *Divertimento* (Piece 6): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants. 108
5.9 Debussy, *La Mer* (Piece 7): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants. 109
5.10 Liszt, *Liebestraum No.3* (Piece 8): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants. 110
5.11 Bach, *Partita No. 2* (Piece 9): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants. 111
5.12 Mean Arousal and Valence (a) and diff. mean Skin Conductance Level and Heart Rate (b) for each piece. White and black dots correspond to the pieces with positive Valence, and black and dark grey dots indicate the pieces with positive Arousal. The numbers indicated next to each dot correspond to the piece ID, as indicated in Table 5.3. 113
5.13 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 subjective feelings of emotion expected to be elicited in the listeners by each piece. 115
5.14 Mean Arousal and Valence (a), and Skin Conductance Response and Heart Rate (b) of the 27 segments of music. White and black dots correspond to the pieces with positive Valence, and black and dark grey dots indicate the pieces with positive Arousal. The label indicated next to each dot correspond to the segments identifiers as indicated in Table 5.4. 117

5.15 Albinoni, *Adagio* (Piece 1): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars. 122

5.16 Grieg, *Peer Gynt Suite No. 1* (Piece 2): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars. 123

5.17 Bach, *Prelude and Fugue No. 15* (Piece 3) : first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars. 124

5.18 Beethoven, *Romance No. 2* (Piece 4): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars. 125

5.19 Chopin, *Nocturne No. 2* (Piece 5): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars. 126

5.20 Mozart, *Divertimento* (Piece 6): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars. 127

5.21 Debussy, *La Mer* (Piece 7): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars. 128

5.22 Liszt, *Liebestraum No.3* (Piece 8): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars. 129

5.23 Bach, *Partita No. 2* (Piece 9): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars. 130

5.24 Linear Discriminant Analysis: only psychoacoustic variables (left); psychoacoustic and physiological variables (right). 133

6.1 Areas covered by the pieces grouped by training (a) and test (b) sets. In both plots, the colours of each cloud of points correspond to all the Arousal/Valence pairs from each piece (see legend). Each point corresponds to the specific location on the 2DES of the Arousal and Valence values on a second by second basis. 142

6.2 Neural network architecture and units identification for Model 1: Input units - sound features (T, Tx, L, P, S and Pv); Hidden units - H_1 to H_5; Memory (context) units - M_1 to M_5; Output units - Arousal (A) and Valence (V). 144

6.3 Neural network architecture and units identification for Model 2. Inputs: sound features (L, T, P, Pv, Tx, S), physiological variables (HR, SCR). Outputs: Arousal (A) and Valence (V). 150

6.4 Arousal and Valence model outputs compared with experimental data for the training data set: Piece 1 (Albinoni, *Adagio*), Piece 2 (Grieg, *Peer Gynt Suite No. 1*) and Piece 3 (Bach, *Prelude and Fugue No. 15*). 154

6.5 Arousal and Valence model outputs compared with experimental data for the training data set: Piece 4 (Beethoven, *Romance No. 2*), Piece 5 (Chopin, *Nocturne No. 2*) and Piece 6 (Mozart, *Divertimento*). 155

6.6 Arousal and Valence model outputs compared with experimental data for the training data set: Piece 7 (Debussy, *La Mer*), Piece 8 (Liszt, *Liebestraum No.3*) and Piece 9 (Bach, *Partita No. 2*). . . . 156

6.7 Neural network detailed architecture including all connections between processing units. 157

6.8 Neural network weight connection matrices: memory to hidden layers (top), input to hidden layers (middle) and hidden to output layers (bottom). The weights are represented as rectangles of variable size and colour: the size is proportional to the weight value and the colour represents the signal of the weight (red for negative and green for positive). 158

6.9 Model weight matrices analysis: each learned weight in the model was removed (value set to 0.0), one at a time. The model performance was measured using the rms error and the values are indicated for each weight removed. Each cell corresponds to the removal of one connection linking two processing units. For an easier reading the rms errors are represented using a colour code: black for those weights that had a small or no effects on the model performance ($rms < 0.09$); for higher errors between gray ($0.09 < rms < 0.30$) and white were used ($rms >= 0.30$). 159

6.10 Qualitative representation of individual relationships between music variables, hear rate and emotion appraisals: summary of observations from model analysis. The direction of the arrows indicates an increase in the variable indicated (the arrow sizes and angles formed with both axis are merely qualitative, and cannot be interpreted in mathematical terms) 164

A.1 Call for participants: this information was distributed in printed and electronic formats. 178

F.1 Sound features extracted with Psysound 3: T. Albinoni - Adagio (piece 1) 192
F.2 Sound features extracted with Psysound 3: E. Grieg - Peer Gynt Suite No. 1 - IV. "In the Hall of the Mountain King" (piece 2) 193
F.3 Sound features extracted with Psysound 3: J. S. Bach - Prelude and Fugue No. 15 - I. "Prelude" (piece 3) 194
F.4 Sound features extracted with Psysound 3: L. V. Beethoven - Romance No. 2 (piece 4) 195
F.5 Sound features extracted with Psysound 3: F. Chopin - Nocturne No. 2 (piece 5) 196
F.6 Sound features extracted with Psysound 3: W. A. Mozart - Divertimento - II. "Allegro di molto" (piece 6) 197
F.7 Sound features extracted with Psysound 3: C. Debussy - La Mer - II. "Jeux de vagues" (piece 7) 198
F.8 Sound features extracted with Psysound 3: F. Liszt - Liebestraum No.3 (piece 8) 199
F.9 Sound features extracted with Psysound 3: J. S. Bach - Partita No. 2 - "Chaconne" (piece 9) 200

ACKNOWLEDGEMENTS.

First of all I would like to express my gratitude to the Portuguese Foundation for Science and Technology (FCT/Portugal) that afforded me an invaluable learning opportunity by providing me with the financial support throughout the duration of this research. I would also like to thank the support of my supervision team during this time. Eduardo Miranda, who gave me the opportunity to start my research in music and to develop the initial stages of this project, and particularly Angelo Cangelosi who patiently guided me through the research process and taught me the fundamental aspects of research and academic life. Also my gratitude goes to Guido Bugmann for his friendship and timely advice. I am also obliged to those who taught, helped and guided me before I started this project. My special thanks to the support of Adriano Carvalho and Luis Rocha during my initial steps in research, and to the people I encountered during that path, especially the generous Henrique Pereira, who is certainly the person that triggered my enthusiasm and love for research.

I have had the good fortune of meeting wonderful people during the course of this project. In their own way, each gave me the opportunity to become aware of the many subtleties of a wide range of subjects and to share the pleasures of discovering and experimenting with new things. In no particular order: - My colleagues and friends with whom I started my research at the Future Music Lab (Peter Beyls, Nikolas Valsamakis, Marcelo Gimenes, Hilary Mullaney, Jaime Serquera, Adolfo Maia, Qijun Zhang, Andrew Brouse, Vadim Tikhanoff), and with whom I had great fun and learned many, many things. - To all the researchers and friends that helped me during these years with their ideas, critical spirit and invaluable conversations (Andrew Hennell, Huck Turner, Oliver Grewe, Marco Mirolli; John Schureman and Denise Carson, whom also edited parts of this manuscript). - A special mention to Leonid Perlovsky whose support was fundamental in important moments of my research.

To my "online" colleagues who were fundamental in several stages of this work. Their experience, ideas and support allowed me to access a wider range of resources, which were used in the course of this research (Stephan Jose Hanson, Gianluca Massera, Sam Ferguson, Andreia Dionisio, Mark Korhonen, and many others). - To the University of Plymouth staff for guiding and supporting us through the academic life, and particularly to Carole Watson and Sue Kendall. - To the reviewers and editors that bring us the fundamental feedback that allows a research project to move forward and to improve it in many aspects, by sharing

their knowledge and dedication to research. - To all the people that participated in my experiments. Some of the most gratifying moments of my work were to share their enthusiasm for music.

I am especially grateful to my family for their guidance and care throughout the years. They afforded me with opportunities to pursue my goals and gave me unconditional love and support. To my good friends João Martins and Patricio da Silva. Their knowledge and care were fundamental in important moments of my research and personal life. To all my friends that throughout the years had an important role in some of my choices in different moments of my evolution.

I want to dedicate this thesis to my wife Rosa, for her companionship and patience, curiosity and enthusiasm, but essentially for being so wonderful and making my days much better. This thesis is also dedicated to the memory of my godfather Manuel Eduardo, whom I never had the opportunity to thank for his kindness and advice in very important moments of my development.

AUTHOR'S DECLARATION.

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

Relevant scientific seminars and conferences were regularly attended at which work was often presented.

Publications :

Coutinho, E. & Cangelosi, A. (in press). The use of spatiotemporal connectionist models in psychological studies of musical emotions. *Music Perception.*

Coutinho, E. & Cangelosi, A. (2008). Psycho-physiological patterns of musical emotions and their relation with music structure. *In Miyazaki, K., Hiraga, Y., Adachi, M., Nakajima, Y. & Tsuzaki, M. (Eds.), Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC10).* Sapporo (Japan).

Coutinho E. & Cangelosi A. (2007). Modeling emotion and embodiment in multi-agent systems. *In H. Hexmoor & C. Thompson (Eds.), Proceedings of 2007 International Conference on Integration of Knowledge Intensive Multi-Agent Systems (KIMAS07).* Waltham (MA, USA): IEEE Press, pp. 133-138.

Coutinho, E. & Cangelosi, A. (2006). The dynamics of music perception and emotional experience: a connectionist model. *In Baroni, R., Addessi, A. & Costa, M. (Eds.), Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC9).* Bologna, Italy: Bologna University Press.

Coutinho, E., Miranda, E. & Silva, P. (2005). Evolving emotional behaviour for expressive performance of music.*In Panayiotopoulos, T., Gratch, J., Aylett, R., Ballin, D., Olivier, P. & Rist, T. (Eds.), Intelligent Virtual Agents: Proceedings of the 5th International Working Conference (IVA2005)* . Berlin (Germany): Springer-Verlag (LNAI 3661), pp. 147-147.

Coutinho, E., Miranda, E. & Cangelosi, A. (2005). Towards a Model for Embodied Emotions. *In Bento, C., Cardoso, A. & Dias, G. (Eds.), Proceedings of the Portuguese Conference on Artificial Intelligence (EPIA2005).* Covilhã (Portugal): IEEE Press, pp. 54-63.

Coutinho, E., Gimenes, M., Martins, J., & Miranda, E. (2005). Computational musicology: An artificial life approach. *In Bento, C., Cardoso, A. & Dias, G. (Eds.), Proceedings of the Portuguese Conference on Artificial Intelligence.* Covilhã (Portugal): IEEE Press, pp. 85-93 .

Coutinho, E., Miranda, E., & Cangelosi, A. (2005). Artificial emotion - simulating affective behaviour. *In Proceedings of the Post-cognitivist Psychology Conference.* Glasgow, Scotland.

Seminars :

A Neural Network Model for the Prediction of Musical Emotions, CogSys Doctoral Consortium. Munich, Germany (June, 2008);

Emotion and Embodiment in Cognitive Agents: from Instincts to Music, International Conference on Integration of Knowledge Intensive Multi-Agent Systems (KIMAS07). Waltham (MA), USA (May, 2007);

The dynamics of music perception and emotional experience: a connectionist model, 9th International Conference on Music Perception and Cognition. Bologna, Italy (August, 2006);

Evolving Emotional Behaviour for Expressive Performance of Music, IVA 2005, Kos, Greece, (September, 2005);

Computational Musicology: An Artificial Life Approach, Artificial Life and Evolutionary Algorithms Workshop @ EPIA05, Covilh, Portugal (December, 2005);

Towards a Model for Embodied Emotions, Affective Computing Workshop @ EPIA05, Covilh, Portugal (December, 2005);

Body Driven Music Performance, Digital Music Research Network Workshop, London, UK (December, 2005);

An Artificial Life Approach to the Evolution of Music, Computer Music Research Seminar Series, Plymouth, UK (March, 2005);

Artificial Emotion: simulating affective behaviour, Adaptive Behaviour and Cognition Research Group seminar, Plymouth, UK (July, 2005);

Further studies :

Composing with Computers I, Massachusetts Institute of Technology (USA), 2007

Composing with Computers II, Massachusetts Institute of Technology (USA), 2007

Fundamentals of Music, Massachusetts Institute of Technology (USA), 2007

Workshop on Algorithmic Computer Music, University of California Santa Cruz (USA), 2006

Certificate of Professional Development in Teaching, Learning and Assessment (partial fulfillment for the UK Accreditation for Teaching in Higher Education), University of Plymouth (UK), 2005.

Grants and Awards :

Honourable Mention from the New York Fellowship Council (Van Alen Institute, New York, USA): Polymnia - the Landscapes of Urban Sonic Life, 2008.

Research grant from the Portuguese Foundation for Science and Technology, 2004-2008

Word count for the main body of this thesis: 39366

Signed:

Date: 16.04.2009

Chapter 1

Introduction

Throughout antiquity the relationship between music and emotion has been acknowledged as a fascinating quality of the human experience. This association is so profound that music is often claimed to be the "language of emotions" and a compelling means by which we appreciate the richness of our affective human life. Music gives a "voice" to the inner world of emotions and feelings, which are often very hard to communicate in words (Langer, 1942).

Despite this long-term interest in the relationships between music and human emotions, their nature is still largely unknown (Panksepp & Bernatzky, 2002). The conceptual difficulties in defining emotions and studying them experimentally can partially justify this lack of information. However, the fact that emotions were neglected by the different fields that purport to study emotions and the psychology of music, have also contributed greatly to the slow pace of this research (Juslin & Sloboda, 2001). It was only after the strong revival of studies on human emotions during the late 19*th* century, that a growing scientific interest in music and its relationships with emotional systems has emerged. Even so, the first comprehensive publications focusing on the relationships between music and emotions, bringing together the main perspectives of contemporary studies, was published as recently as 2001 (Juslin & Sloboda, 2001). Only now is music taking

its rightful place at the very centre of scientific interest. One of the main reasons is that music is finally being recognised as a key factor in our understanding of the mechanisms of the human emotional systems (Koelsch, 2005; Gaver & Mandler, 1987).

For most of the last century, the view generally held on musical emotions emphasised their individual and cultural aspects rather than the affective systems supporting them (Panksepp & Bernatzky, 2002). In Meyer's words: "If we then ask what distinguishes non-emotional states from emotional ones, it is clear that the difference does not lie in the stimulus alone. The same stimulus may excite emotion in one person but not in another. Nor does the difference lie in the responding individual. The same individual may respond emotionally to a given stimulus in one situation but not in another. The difference lies in the relationship between the stimulus and the responding individual" (Meyer, 1956, page 11).

Such a perspective has emphasised the strong belief that musical emotions are a highly individual and culture-dependant experience, thus delivering the central role in the musical experience to the learned cultural aspects. Paradoxically, a considerable corpus of literature, regarding the effects of music on human emotions, consistently reports that listeners often agree rather strongly about the emotion expressed by a particular piece of music (or even in particular moments or sections; see Juslin & Sloboda, 2001). This suggests that the same music stimulus can induce similar affective experiences in different listeners, occurring, to some extent, independently and consistently across individual, situational and cultural contexts.

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. Nevertheless, the affective power of sound and music suggests that it may be related to the deeper affective roots of the human brain (Zatorre,

2005; Panksepp & Bernatzky, 2002; Krumhansl, 1997). It is as if music stimuli encoded and symbolised certain affective features of our brain systems, possibly generated by lower subcortical regions (Blood, Zatorre, Bermudez, & Evans, 1999; Blood & Zatorre, 2001) where certain affective states are organised (Damasio, 2000). Moreover, evidence that subcortical mediation is involved in the process of emotional responses to music (Blood & Zatorre, 2001; Blood et al., 1999), suggests that cognitive attributions may not be essential for music to elicit emotions in the listener. This reinforces the importance of understanding the biological affective roots of musical sounds (Panksepp & Bernatzky, 2002).

Certain basic neurological mechanisms related to motivation / cognition / emotion automatically elicit a natural response to music in the receptive listener. This gives rise to profound changes in the body and brain dynamics, and to the interference with ongoing mental and bodily processes (Panksepp & Bernatzky, 2002; Patel & Balaban, 2000). This multimodal integration of musical and information takes place in the brain (Koelsch, Fritz, Cramon, Müller, & Friederici, 2006), suggesting the existence of complex relationships between the dynamics of musical emotion and the perception-action cycle response to musical structure. This is also the belief of several researchers (e.g. Gabrielsson & Lindström, 2001), who postulate that the way musical elements are organised in time is linked with emotional responses in the listener, thus implying the existence of a causal, underlying relationship between musical features and emotional response. At the very least, it is plausible to assume the existence of certain neural networks that deal with the patterns of sound. New conceptualisations of the "emotional brain" (Damasio, 2000) support this possibility. Recent neuroscientific studies on patients with impaired musical cognitions, who still maintain their affective experiences with music, (Peretz, 2001; Peretz, Gagnon, & Bouchard, 1998) have provided evidence for such a claim. Some of the emotional effects of music appear not to be cognitively

mediated, which points to and is consistent with the capacity of music to elicit the desire for bodily movements and to induce physiological arousal (Grewe, Nagel, Kopiez, & Altenmuller, 2007; Guhn, Hamm, & Zentner, 2007; Khalfa, Peretz, Blondin, & Manon, 2002; Panksepp, 1995; Iwanaga & Tsukamoto, 1997; Krumhansl, 1997; Witvliet & Vrana, 1996; Vanderark & Ely, 1993).

All these findings provide evidence of the universality of musical affect and that cognitive mediation is not a required element in music appreciation. In this way, for the affective experience to happen, it is plausible to think that the listener derives affective meaning from the nature of the stimulus. Some theories (Panksepp & Bernatzky, 2002; Clynes, 1978; Janata & Grafton, 2003) suggest that music derives its affective power from dynamic aspects of the brain systems. Support for this research comes from studies, with brain damaged patients, which show that the emotional appreciation of music can be maintained even in the presence of severe perceptual and memorisation deficits, thus reinforcing the idea that subcortical mediation is involved in "emotional judgements" (Blood et al., 1999; Blood & Zatorre, 2001).

1.1 Scope of research

By focusing on the music stimulus and its features as important dynamic characteristics of affective experiences, many studies suggested the influence of various structural factors on emotional expression (e.g. tempo, rhythm, dynamics, timbre, mode, harmony, among others; see Gabrielsson & Lindström, 2001 and Schubert, 1999a for a review) . Unfortunately, the nature of these relationships is complex, and it is common to find rather vague and contradictory descriptions, especially when the music structural factors are considered in isolation (Gabrielsson & Lindström, 2001).

This thesis presents a novel methodology to analyse the dynamics of

emotional responses to music. The approach consists of a computational investigation of musical emotions based on spatiotemporal neural networks sensitive to structural aspects of music. The computational studies are backed up by experimental data, such as the models that are trained on human data to "mimic" human affective responses to music and predict new ones. The affective responses, as considered in this thesis, are limited to subjective feelings of emotion. The dynamics of emotional responses to music are investigated here in terms of computational representations of perceptual processes (psychoacoustic features) and self-perception of physiological activation (peripheral feedback).

1.2 Organisation of this thesis

The first two chapters, following this introduction, will review experimental and computational studies on music and emotion. Chapter 2 presents the theoretical and experimental background for this thesis. First, human emotions are discussed at the neurobiological level. Then, a description of the main mechanisms of emotion serves as an evaluation of the different mechanisms by which music can convey emotion to the listener, and how they can be quantified. This chapter includes an overview of the most common methodologies used to quantify musical emotions, with a special emphasis on methods based on the subjective feelings of the listener and upon physiological activation.

It will be shown that the complexity of experimental data on music and emotion studies, require capable methods of analysis, which allow for the extraction of relevant information from experimental data. In Chapter 3 a class of spatiotemporal (Kremer, 2001) connectionist models will be proposed as a capable paradigm to analyse the interaction between sound features and the dynamics of emotional ratings. The proposal of using computational models for this study aims to investigate the relationships between music structure and

emotion in more detail and includes their computational (abstract) representation.

In the following three chapters I will present new computational and experimental studies. Chapter 4 describes a new neural network model of musical emotions, which accounts for the subjective feeling component of emotions. The experimental data for this simulation experiment was obtained from a study conducted by Korhonen (2004a)[1]. The neural network is trained to predict affective responses to music, based on a set of psychoacoustic components extracted from the music stimuli. An analysis of the network dynamics provides new information about the relationships between the sound and its affective dimensions.

In Chapter 5 a new experimental study is presented which is based on the continuous response methodology (Schubert, 1999b) to obtain a listeners affective experience with music. Participants were asked to report their subjective feelings while listening to music, while at the same time, their heart rate and skin conductance levels were recorded. Evidence provided in this study suggests the existence of relevant interactions between the psychological and physiological components of emotion.

This data is used in Chapter 6 to develop a new computational investigation and for the extension of the neural network model presented in Chapter 4, to include physiological cues. The aim of this chapter is to verify if physiological activity has meaningful spatiotemporal relationships with the affective response.

The last chapter of this dissertation, Chapter 7, summarises the research, and discusses its implications and contributions to the field of music perception, emotion and cognition. A set of recommendations is also included which discuses some of the potential extensions and applications of this model.

[1]Data available online at `http://www.sauna.org/kiulu/emotion.html`.

Chapter 2

The nature of emotion in music: Theoretical and experimental investigations

As a background to this thesis, this chapter presents a literature review of the state of the art research and ideas on musical emotions. Due to the multimodal nature of emotions, modern research is focusing on particular components of its modalities and experiential counterparts (such as feeling states or other manifestations). Initially, the neurobiological framework for emotions theorised by Damasio (1994, 2000) will be used as the substrate for a conceptualisation of emotions, their multimodal nature and classes of inducers. As it will be discussed, emotions arise as a construct of complex psychological and physiological processes. An overview of these mechanisms is given in this chapter, as well as an examination of their potential involvement in musical emotions, and the focus on auditory stimuli as potential emotion-competent-stimuli (Damasio, 2000). The discussion on whether music induces emotions, or merely represents them, will also be discussed in the light of "emotivist" and "cognitivist" perspectives on musical emotions (Kivy, 1989). The last part of this chapter, discusses the main

methods to measure emotions experimentally in music research.

2.1 A substrate for emotions

In the very influential book "The Feeling of What Happens: Body, Emotion and the Making of Consciousness", Damasio (2000) proposed a neurobiological framework in which he demonstrates that certain organisational principles in the brain might reflect emotional states. He attributes to emotions a major role in the general economy of the mind, by suggesting that emotions and feelings are part of the neural machinery for biological regulation (whose core is formed by homeostatic controls, drives and instincts). Emotions are also inseparable from the idea of reward or punishment, of pleasure or pain, of approach or withdrawal, of personal advantage or disadvantage.

Within this framework, emotions are complicated collections of organised chemical and neural responses with some regulatory role to play, leading in one way or another to the creation of circumstances advantageous to the organism. The biological functionality of emotions include the production of a specific reaction to the inducing situation (e.g. run away in the presence of danger), and the regulation of the internal state of the organism such that it can be prepared for the specific reaction (e.g. increased blood flow to the arteries in the legs so that muscles receive extra oxygen and glucose, in order to escape faster). Damasio (2000, pp. 68-69) summarises the sequence of events in the process of emotion in the following manner:

1. "engagement of the organism by an inducer of emotion" (e.g. a particular object processed visually);

2. "the signals consequent to the processing of the object's image activate all the neural sites that are prepared to respond to the particular class

of inducer to which the object belongs". These sites have been "preset innately", although past experience has modulated the manner in which they are likely to respond;

3. "emotion induction sites trigger a number of signals toward other sites (for instance, monoamine nuclei, somato-sensory cortices, cortices) and toward the body (for instance, viscera, glands)".

Damasio categorises emotions into three categories related to their "biological functionality": *background*, *primary* (or *basic*) and *secondary* (or *social*). *Background* emotions are defined as certain responsive conditions of the internal state engendered by ongoing physiological processes, or by the organisms' interactions with the environment, or both. These emotions endow us with, among others, the background feelings of tension or relaxation, of fatigue or energy, of wellbeing or malaise, of anticipation or dread. They target more internal rather than external processes. They are also richly expressed in musculoskeletal changes, for instance through subtle body postures and overall shaping of body movement. *Primary* emotions (fear, anger, disgust, surprise, sadness, and happiness) are considered to be what other studies call the basic emotions, a set of shared predispositions that allows us to respond in a more or less stereotyped way when certain features of stimuli are perceived. The limbic system is often associated with this process, as a unit involved with the detection of these features, and the source of certain brain signals that alter current cognitive activity and activate certain biological responses.

On the top of these two groups appear the *secondary* or *social* emotions (sympathy, embarrassment, shame, guilt, pride, jealousy, envy, gratitude, admiration, indignation, and contempt). Rather than predispositions these constitute "learned" connections related to the state of the organism and the stimuli that triggered the emotional process, both at conscious and unconscious levels. From this category of emotion Damasio derives his "somatic markers

hypothesis" (Damasio, 1994), a refinement of the mind to assess and respond quickly to a set of conditions of the internal and external environments of an organism, that are associated with the detection of an "emotional-competent-stimulus". In situations that require decisions involving complex and conflicting alternatives, cognitive processes may become overloaded, and the somatic markers can aid the decision process. Such associations are reinstated physiologically and drive cognitive processing when needed towards the selection of the appropriate action.

2.2 The modalities of Emotion

A crucial issue in studies of emotion is the conceptualisation of the processes that underlie the elicitation and differentiation of emotional responses. The complexity and diversity of emotions and its mechanisms have promoted different approaches to its study, giving rise to several theories and definitions emphasising its multiplicity of mechanisms and interactions. These approaches frequently involve the study of emotion from its different elements, frequently divided into three main classes: subjective feelings, behavioural changes (motor expression and action tendencies) and physiological arousal (Oatley & Jenkins, 1996; Scherer & Zentner, 2001).

A concept that will be used throughout this chapter is appraisal. This is the core concept for cognitive theories of emotion (Arnold, 1960). Appraisal accommodates the process of evaluation of a stimulus (a fixed sequence of stimulus evaluation checks (Scherer, 1999)) 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. Further developments and refinements on cognitive

theories of emotion followed Arnold's view. For Lazarus (1991), cognition is the basis for differentiation among emotions and for the process of coping with the specific situation in which an organism is involved, which accommodates direct actions (including behavioural changes) and cognitive reappraisal processes. The corollary of this theory is that appraisals are both necessary and sufficient for emotion. In the following sections, the way appraisal theories accommodate the different modalities of emotion is also discussed.

2.2.1 Subjective feeling

Damasio (2000) defines Feelings as the mental representations arising from the neural patterns that represent an organisms' internal changes which follow and characterise an emotion (e.g. interactions with emotion eliciting objects or states, which can be either internal or external). Feeling states were also believed to modulate functions such as decision making and interpersonal interactions. The subjective feeling of emotion refers to its experienced qualities, based on the understanding of the felt experiences from the perspective of an individual. If the perception of emotional events leads to rapid (some automatic and stereotyped) emotional responses, feeling states have a slower modulatory effect in cognition (and ultimately behaviour and decision making, according to the nature and relevance of the eliciting stimulus). More generally the feeling of emotion can be seen as a (more or less diffuse) representation that indexes all the main changes (in the respective components) during an emotional experience. This is the compound result of event appraisal, motivational changes, and proprioceptive feedback (from motor expression and physiological reactions) (Scherer, 2004). These conscious "sensations" are an irreducible quality of emotion, unique to the specific internal and external contexts, and to a particular individual (Frijda, 1986; Lazarus, 1991).

2.2.2 Behavioural changes

Two of the most important classes of "emotional behaviour" are motor expression (expressive cues) and action tendencies (e.g. Ekman, 1973; Frijda, 1986). The first type, expressive cues, is often associated with a Darwinian perspective on emotion. Charles Darwin, in his 1872 book "The Expression of Emotion in Man and Animals", presented emotions as reaction patterns shaped through the evolutionary process: responses that had a survival value were kept. These emotions are considered to be common to the human species (and some also shared with other mammalians), serving biological and/or communicative functions, and referred to as basic (or universal) emotions. One of the most relevant findings of this ideological tradition is the universality of facial expressions (Ekman, 1973, 1999). In a cross-cultural study Ekman showed that at least six emotions (happiness, sadness, fear, anger, disgust and surprise) are expressed through facial expressions (different muskeletal configurations) recognised cross-culturally. Expressive cues are also in the base of studies on the social implications and facets of emotion (e.g. Izard, Huebner, Risser, McGinnes, & Dougherty, 1980), whereby emotions are seen as socially acquired constructed patterns which are culturally shared and fulfil a social purpose. Both emotion and the subjective experience are considered to be culturally constructed. These theories recognise the existence of biological foundations for emotions (like Ekman's facial expressions), but place them at a lower hierarchal level, secondary to the socially constructed mechanisms.

Focusing on the process that follows appraisal, Frijda (1986) developed the concept of emotion as a state of action readiness (or action tendency), a type of motivational state. Different emotions are mediated by an appraisal of the situation in relation to an individual's current state, past experience, beliefs, goals, from which follow certain patterns of behaviour and the feeling stage and the awareness of the emotional experience. These components are interactive

and mutually influential. Dispositions are sensitive to specific appraisal patterns, which then modulate arousal (physiology) and motivation (action readiness), facilitating behavioural changes, with or without effective consequences. These behaviours can be suppressed or hidden in favour of other competing mind processes. Scherer (1999) emphasises the sequential nature of the appraisal process. Appraisal consists of a sequence of stimulus evaluation checks, which facilitate the differentiation of different emotions. This sequence is considered to be invariant within the recursive process that constitutes appraisal. Scherer and his colleagues have developed elaborate appraisal theories and created several models that use the sequence hypothesis as its core, taking a functional approach to emotion.

2.2.3 Physiological arousal

Arousal is the term that defines a state of "alertness". Physiological arousal refers to the level of physiological activation that characterises that state. Increased arousal is associated with increased heart rate, increased body temperature, increased respiration rate (increased oxygen consumption), and many other physiological changes. From a neurophysiological perspective, it involves the activation of the autonomic nervous system (ANS), the reticular activating system (in the brain stem), and the endocrine system. It leads to a condition of "sensory alertness" and readiness to respond.

In emotion studies, physiological patterns of activation gained special attention after the work of William James (1884) and Carl Lange (see Cannon, 1927). What is now known as the James-Lange theory, posits that it is through the self-perception of bodily changes that the emotion arises. The bodily changes are considered to occur as a reaction to a certain stimulus, and emotions and feelings arise from the perception of these changes. The corollary of this theory is that without the perception of the body, emotion cannot take place. Although

a limitation in terms of the definition of emotion, these ideas were followed by important theoretical and experimental studies that unveiled the importance of the body states in the process of emotion. Important theoretical developments were put forward by Cannon (1927), who supports the idea that physiological activation is a component of the emotional experience. Unlike James, Cannon claims that physiology plays a role in the emotion construct, but doesn't consider it to be the emotion itself.

Currently neurobiological models of emotion recognise the importance of higher neural systems on visceral activity (top-down influences) but also the 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 somatovisceral substrate, the bottom-up ones are suggested to serve to bias emotion and cognition towards a desired state (e.g. guiding behavioural choice, Bechara, Damasio, & Damasio, 2003).

Modern conceptualisations propose that a stimulus (appraised via cortical or subcortical routes) triggers physiological changes, which in turn facilitates action and expressive behaviour. In this way, together with other components of emotion, physiological activation contributes to the affective feeling state. Implicitly individuals may use their body state as a clue to the valence and intensity of the emotion they feel (Dibben, 2004). As reviewed by Philippot, Chapelle, and Blairy (2002), three main approaches to the integration of visceral activity into models of emotion can be distinguished: (i) the undifferentiated arousal model, (ii) the cognitive appraisal model, and (iii) the central network model. The main idea behind the first model (Reisenzein, 1983; Schachter, 1964) is that body responses increase with emotional intensity, but their pattern is not differentiated across the different emotional states. In this line cognitive information and/or the specific context differentiate the type of emotion, while

bodily activation (arousal) determines the intensity of that emotion. One practical prediction of this model is that the perception of the emotional intensity can be influenced by the arousal intensity. The main finding of this research (see Reisenzein, 1983) has been the fact that after exposure to an arousing stimulus, the following emotional feeling state is intensified. This phenomenon is called activation transfer (Zillmann, 1983).

The second model focuses on the body changes as a function of the cognitive appraisal processes (Scherer, 1984), or its direct output: action readiness (Frijda, 1986). From this perspective patterns of body activation simply follow the result of appraisal, or more precisely the combined result of the several cognitive appraisal components. The fact that the body itself might generate emotional states is quite marginalised in this model. Finally, from the third model perspective, emotion is seen as a construct of different components, which are centrally organised by different neural or cognitive networks. Some researchers suggest that these pathways are innate (e.g. Ekman, 1999; Tomkins, 1980; Izard, 1979), while others consider them to be cognitive schemata developed as a function of an individuals experience (e.g. Lang, 1979; Philippot et al., 2002). The patterns of body changes are considered to be differentiable across emotions and events appraised (via cortical or subcortical routes) can elicit physiological changes that facilitate action and expressive behaviour (as in the previous model). The central network model also conceives that a specific emotion can be elicited by creating specific body state patterns (peripheral feedback), even outside the awareness of the individual (Damasio, 1994).

Researchers studying facial expression and subjective feelings reported empirical support for the peripheral feedback hypothesis. A consistent amount of evidence was presented which suggests that the manipulation of facial expression can affect the emotion state of an individual (e.g. McIntosh, 1996; Manstead, 1988). Philippot et al. (2002) have shown that respiratory changes

are subjectively differentiated across emotions, and that respiration manipulation can also induce emotional states in listeners outside their awareness.

2.3 The nature of Emotion in Music

The ability of music to stir human emotions is a well known fact (Gabrielsson & Lindström, 2001). However, the manner in which music contributes to those experiences remains obscured. One of the main reasons is the large number of syndromes that characterise these experiences. Another obstacle is the subjective nature that characterises emotion. For instance, the emotion created by a piece of music may be affected by memories, the environment and other situational aspects, the mood of the person listening, 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).

Despite the number and complexity of factors associated with musical emotions, a considerable corpus of literature on the emotional effects of music in humans 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 be found in Juslin and Sloboda (2001); see also Gabrielsson and Lindström (2001)). Naturally this leads to a question: can the same music stimulus induce similar affective experiences in all listeners, somehow independently of acculturation, context, personal bias or preferences?

2.3.1 "Emotivist" vs. "Cognitivist"

There are two main complementary views on the relationships between music and emotions. "Cognitivists" claim that music simply expresses emotions that

the listener can identify, while "emotivists" posit that music can elicit affective responses in the listener (Krumhansl, 1997; Kivy, 1990). In contemporary research these separate views are also closely related to the distinction between perception and production of emotion, although only a few studies have addressed this question directly. Instead some researchers (e.g. Dibben, 2004; Gabrielsson, 2002; Krumhansl, 1997; Witvliet & Vrana, 1996) used self report of subjective feelings and physiological measurements to argue that listeners do indeed feel emotions when they listen to music. Others, e.g. Scherer and Zentner (2001), suggest that emotions induced through music may be a different subset from the emotions that music can express. In another study, Zentner, Meylan, and Scherer (2000) concluded that listeners can both feel and perceive emotions in music, but that they may vary according to the music genre. Gabrielsson (2002) also suggests that there is evidence for a positive relationship between perception and induction of emotion. However, Gabrielsson argues that this belief also reflects an assumption that listeners responses are determined primarily by musical factors. If the role of personal and situational factors is acknowledged, this positive relationship becomes less likely to exist. Perceived emotions and felt emotions differ, but the latter seems to be the most important reason for listening to music (Sloboda, 1991).

One of the most influential works from a "cognitivist" perspective on musical emotions was the book *Emotion and meaning in Music* by Meyer (1956). He developed a theory in which musical emotions depend mainly upon expectations about the unfolding events and their meanings, which create patterns of tension and release in the listener (Meyer, 1956). For Meyer, expectation is a necessary condition for emotion and meaning to be conveyed in music. The nature of these expectations derives from the development of psychological schemas of more or less complex systems of sound relationships. These include the general Gestalt principles for perceptual organisation, but mainly psychological

schemas derived from the interaction with a given (musical) culture. Without the "stylistic experience" music becomes meaningless and consequently lacks in affect. Empirical support for Meyer's ideas has come from different formalisations of his theory (e.g. Narmour, 1992; Krumhansl, 1991; Cuddy & Lunney, 1995). Meyer's "cognitivist" perspective is especially evident in a passage of his 1956 book: "... when a listener reports that he felt this or that emotion, he is describing the emotion which he believes the passage is supposed to indicate, not anything which he himself has experienced" (Meyer, 1956, p. 8).

"Emotivists" claim that music can itself elicit emotions in listeners. Although our affective experiences with music are ultimately individual and culturally dependent, there are certain musical dimensions and qualities which induce similar affective experiences in all listeners, across all cultures, and independent of context and personal biases or preferences. Evidence supporting such possibility comes from cross-cultural studies. Balkwill & Thompson's (1999) has shown western listeners (with no familiarity with North Indian *ragas*) listened to *Hindustani* music and were able to identify emotions of joy, sadness, and peace. Additionally, some of the most compelling observations, have been obtained in neuroscientific studies in the context of severe deficits in music processing after brain damage and recurring to modern brain imaging techniques, which are now revealing that the emotional impact of music is substantially dependent on both direct and indirect effects on subcortical emotional areas of the human brain. For instance, the suggestion that music itself can elicit emotions without the involvement of cognition, favouring the "emotivist" view on musical emotions, can be found in Peretz et al., 1998. Peretz et al. (1998) described a patient (I.R.) suffering from severe loss of music recognition and expressive abilities. I.R. showed no evidence of impairment in the auditory system, but still she could not discriminate pitch and temporal deviations in the music. Even violations of the scale structure, or judgements of adequacy of a pitch as the ending of

an harmonic sequence (tonal closure), were impossible to I.R. to discriminate. Despite all this, I.R. still claims the capacity to enjoy music. In the experiment, the patient was able to derive the emotional tone of the excerpts, manipulated in terms of tempo and mode to achieve the intended emotional qualities. Although I.R. was not aware of the music manipulations[1], she performed as well as the control group on the affective content identification task. This study shows that the perceptual analysis of the music input can be maintained for emotional purposes, even if impaired for cognitive ones. Peretz et al. (1998) suggest the possibility that emotional and non-emotional judgements are the products of distinct neurological pathways.

The advent of brain imaging techniques has allowed for an increasing specification of the neural underpinnings of music processing and different concomitants with the listening experience, including aesthetic and emotional responses. Nevertheless, the fact that music interacts with a wide range of neural networks with possible access to emotion related areas suggests that there is no restricted brain "module" exclusively devoted to musical appreciation. Our current understanding is that emotional circuits are widely distributed in the brain, with roots in subcortical neural networks linked with many areas in cortical regions (Panksepp, 1998a). Accordingly, music is bound to access these emotional systems at many levels. As an example, the auditory processing of musical information can access "emotional" systems through temporal lobe inputs into the amygdala, or the frontal and parietal cortical inputs into other basal ganglia such as the nucleus accumbens, or even more direct inputs to limbic areas such as the cingulate and medial frontal cortices (Blood et al., 1999; Blood & Zatorre, 2001; Panksepp & Bernatzky, 2002). This human complex sensitivity to emotional sounds may be related to survival benefits that subtle emotional communications had for us during our evolutionary history (Panksepp

[1]In the screening tests, used to test I.R.'s ability to process music, she did not give any indication that she could perceive and/or interpret pitch and temporal variations in melodies.

& Bernatzky, 2002). Take as an example the fact that the inferior colliculus (a mandatory brainstem way-station for auditory processing): this is a brain area which mediates affective processes (Bagri, Sandner, & Di Scala, 1991 cited in Panksepp & Bernatzky, 2002) and is richly endowed with opiate receptors (Panksepp & Bishop, 1981), suggesting that the inferior colliculus may mediate attachments we develop to certain sounds (e.g. voices) and to certain types of music. The inferior colliculus is also adjacent to the periaqueductal gray (PAG) an area where all emotional systems converge upon a coherent self-representation of the organism (Panksepp & Bernatzky, 2002).

The involvement of these (and other) brain systems in response such an abstract stimulus as is the case of music, suggests an emergent property of the complexity of human cognition: it may be the case that the formation of anatomical and functional links between phylogenically older and newer systems, may have increased our capacity to derive pleasure from music (Panksepp & Bernatzky, 2002). Some of the pathways involved in music processing were found to involve the activation of emotion, arousal and reward/motivation related areas (e.g. Koelsch et al., 2006; Blood & Zatorre, 2001; Blood et al., 1999), also recruited in response to biologically relevant stimuli such as food and sex, and artificially activated by drugs of abuse[2]. Modern conceptualisations of the emotional brain acknowledge these and other brain circuits as fundamental for the generation of affective responses (e.g. Panksepp, 1998; Damasio, 2000), which can operate even outside an individual's awareness. The ability of certain types of music to evoke a desire for bodily movements, and various autonomic changes, also reinforces the idea of an active involvement of subcortical regions (e.g. Blood & Zatorre, 2001) in response to music. It is nevertheless important, as referred previously, not to ignore the fact that great amount of human musical tendencies and appreciations are undoubtedly learned, and that both perspectives should be

[2]For a detailed discussion on the biological roots of music and its possible implications for evolutionary theories refer to (Huron, 2001)

integrated, as they seem to be highly interactive during the processing of musical stimuli. This has been the perspective adopted by Huron (2006), whom, based on current theories of emotions, psychological data, and recent neuropsychological findings, has revisited Meyer's "Emotion and Meaning in Music" (Meyer, 1956) to draw from evolutionary theories and biological evidence a new theory of musical expectations and their evaluation as a major source of pleasure in music listening.

Modern theories of musical emotions emphasise that music derives its affective power from dynamic aspects of the brain systems. As seen, support for this research comes from studies with brain damaged patients, which show that the emotional appreciation of music can be maintained even in the presence of severe perceptual and memorisation deficits, though reinforcing the idea that subcortical mediation is involved in "emotional judgements" (Blood et al., 1999; Blood & Zatorre, 2001). Due to these interactions certain basic mechanisms related to motivation/emotion in the brain can be elicited by music. This gives rise to the changes in the body and brain dynamics, and to the interference with ongoing mental and bodily processes (Panksepp & Bernatzky, 2002; Patel & Balaban, 2000). This multimodal integration of musical and information might take place in the brain (Koelsch et al., 2006), suggesting the existence of relationships between the dynamics of musical emotion and the perception of musical structure. There is evidence of the universality of musical affect and that cognitive mediation is not a required element in music appreciation and so, for the affective experience to happen, it is plausible to think that the listener derives affective meaning from the nature of the stimulus. This is also the belief of several researchers (e.g. Gabrielsson & Lindström, 2001), who posit that the way musical elements are organised in time is linked with emotional responses in the listener, implying the existence of a causal, underlying relationship between musical features and emotional response.

2.3.2 Music elements and the expression of emotion

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 organise these dimensions according to different 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 (e.g. Levitin, 2006). 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, and so the subjective experience should be, at least partially, dependent on those features (Kellaris & Kent, 1993).

Hevner's early studies (Hevner, 1936) are amongst the first to systematically analyse which musical parameters are related to the reported emotion. Major versus minor modes, firm versus flowing rhythm, the direction of melodic contour, and complex/dissonant versus simple/consonant harmonies, were the music qualities manipulated in each piece presented in the cited experiment. Hevner had subjects listen to these pieces and select an adjective (from "Hevner's adjective circle") that best described the emotional content of each piece. Tempo and mode had the strongest impact on the listener. For example, piano music played fast in major mode was labelled as cheerful while the slow piece in minor mode was labelled as sensitive and "dreamful". Since then, a core interest amongst music psychologists has been the isolation of the perceptible factors in music which may be responsible for the many observed effects, and a fairly regular stream of publications have attempted to clarify this relationship. A summary of these relationships will be presented in the next chapter. For a review of past studies please refer to Gabrielsson and Lindström (2001).

Despite the number of studies that investigated the relationships between music structure and emotional expression, there are still many doubts, uncertainties and contradictory observations. Added to the difficulties in quantifying certain musical dimensions (e.g. rhythm, timbre, harmony, among others), methodologies faced difficulties in manipulating music factors in isolation. Moreover, there are no music factors that work in isolation, meaning that their effect is at least mediated by the relationships with other elements (Gabrielsson & Lindström, 2001; Hevner, 1936; Rigg, 1964)[3].

Another negative aspect of many studies has been the quantification of the musical factors themselves. Many researchers have used qualitative descriptions to describe the perceived changes in the music, often supported by discrete scales considering only two extreme levels (e.g. fast and slow in the case of tempo). Such an approach leaves aside all the intermediate levels of these variables assuming that they behave linearly within the extreme categories defined, thus neglecting a wide range of musical possibilities and complexity of interactions between variables. Another problem of such an approach arises when we consider that music can elicit a wide range of emotions in the listener. A piece of music is characterised by changes over time, which are a fundamental aspect of its expressivity. The dynamical changes over time are perhaps the most important ones (Dowling & Harwood, 1986), especially if we consider that musical emotions may exhibit time-locking to variations in psychological and physiological processes, consistent with a number of studies that show temporal variations in affective responses (e.g. Goldstein, 1980; Nielsen, 1987; Krumhansl, 1997; Schubert, 1999a; Korhonen, 2004a). The static attributes of music are only partially responsible or indicative of emotional response to music, which can be intense and momentary (e.g. Dowling and Harwood (1986)). This approach is

[3]Many of the perceived changes in a piece depend on different features of the sound perceived. For instance, the loudness level depends on the pitch level. The perception of melodic interval depends on its direction, tempo, rhythmic pattern, loudness and pitch levels.

also congruent with the view advocated by Langer (1942) on the existence of expressive forms ("iconic symbols") of emotions in all art forms. She believed that art, and music in particular, are fundamental forms of human physical and mental life.

2.4 Measurements of musical emotions

The majority of studies involving the measurement of human emotions make use of mainly three general classes of quantities (Berlyne, 1974): physiological responses (e.g. heart rate, galvanic response), behavioural changes or behaviour preparation (e.g. facial expressions, body postures) and the subjective feeling component (self report of emotion: e.g. checklists and scales). These reflect the three main modalities of emotion described earlier (see Section 2.2).

The two modalities most often measured in music and emotion research are subjective feelings and physiological arousal. One is associated with the feeling of emotion ("felt" or "perceived") during music listening, and the way the music may trigger or represent a specific emotion or representation of it. The other investigates patterns of physiological activity associated with music quantities (such as the relationships between tempo and heart rate), as well as the peripheral routes for emotion production (how physiological states relate with entire emotional experience). This process is usually referred to as peripheral feedback. Several studies have found evidence that such a phenomenon can be responsible for the generation of everyday emotions (Philippot et al., 2002) but also musical emotions (Dibben, 2004) (described later in this section). Both routes have been shown to be independent but interactive (e.g. Ekman, Levenson, & Friesen, 1983; McIntosh, 1996 and Feldman, 1995).

2.4.1 Subjective feeling measurements

Schubert (1999a) provided a detailed review of the most common self report measures of emotional responses to music. The self report measures are classified in three broad categories: open-ended, checklists and scales. Figure 2.1 shows an adapted version of the original diagram used in Schubert (1999a, p. 34).

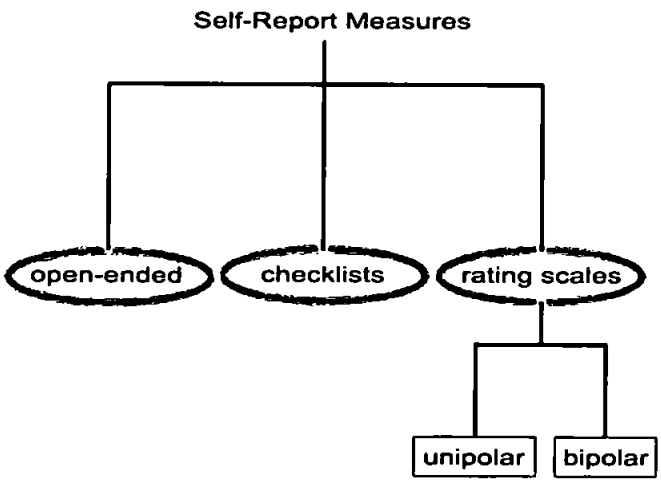


Figure 2.1: Taxonomy of self report measures of emotional response to music. Adapted from Schubert (1999a, p. 34)

Open-ended measures allow the respondents to form their own answer, instead of choosing from a predefined set. This method imposes the minimum restrictions on the listeners, and has been widely used (e.g. Gilman, 1891, 1892; Downey, 1897; Weld, 1912; Lee, 1932; Valentine, 1914; Washburn & Dickinson, 1927; Sherman, 1928; Nelson, 1985; Gabrielsson & Lindström, 1995; Gabrielsson, 1991; Watson, 1942; Flowers, 1983). They are rich measures of musical emotions because they can provide a very detailed level of description of particular and global aspects of the emotional experience. Despite these advantages open-ended measures have serious limitations, including the

frequent occurrence of analytical errors (Rigg, 1964), difficulties of interpretation due to the subjective nature of the report and to the way listeners express themselves (Campbell, 1942; Kratus, 1993), the inconsistencies in encoding the responses and the difficulties created for the listening task. Additionally, it is only after a meaningful reduction of the large amount of data collected, that the open-ended responses are able to be encoded and analysed systematically.

Using a checklist based framework[4], participants are asked to select a word (or several) from a list that best describes their emotional state (or the emotion expressed by the music). The lists of words can be study specific[5] or predefined lists (e.g. basic emotions). Examples of such studies are: Heinlein (1928),Hampton (1945),Gundlach (1935),Hevner (1937),Clynes (1977),Mull (1949) and Sloboda (1991). Compared with open-ended measures, checklists provide better consistency in responses (Rigg, 1937) but, nonetheless, they still preserve several disadvantages. An important one is the lack of established relationships between the verbal categories used and the emotional experience (Valentine, 1962). Other disadvantages are the limited description of the variety and complexity of the emotional experiences with music (Kratus, 1993), the difficulties in creating list of words and phrases covering a sufficient semantic space of the emotional qualities studied (e.g. Russell, 1989) or the limited analysis that can be performed on the data (Schubert, 1999a).

The third type of measures shown in Figure 2.1 are rating scales. This is the most common method used for the measurement of emotion in psychological experiments. A scale permits the quantification of a particular variable into discrete or continuous values. The scales generally[6] used are either unipolar

[4]The most commonly used method during the first half of the 20th century.

[5]Some researchers (Flowers, 1988; Sherman, 1928; Valentine, 1962) suggest that open-ended measures provide a good source of information for the creation of lists of words and phrases for more controlled responses.

[6]Matching or ranking strategies, as described by Schubert (1999a), are not included in this thesis because they cannot be clearly distinguished from the other types of measures described. As an example, using matching measures, the word or sentence is chosen from a list to describe a specific stimulus (just like with a checklist), except that then, the chosen qualifier is removed from

(measures only a single dimension per scale) or bipolar (measures two dimensions per scale). An example of the application of these scales is to ask listeners how happy a piece of music was, offering a discrete scale varying within a determined range (e.g. 1 to 5) whose labels attached to the extreme values are *not happy* and *very happy* (unipolar rating scale); if, instead, the scale has the opposite meanings attached to its extreme values (e.g. *very sad* and *very happy*), the rating scale becomes bipolar. Examples of the use of both types of rating scales are: Wedin, 1969; Asmus, 1985; Thayer, 1986; Gabrielsson & Juslin, 1996 (noncontinuous unipolar); Goldstein, 1980; Panksepp, 1995; Waterman, 1996 (continuous unipolar); Gray & Wheeler, 1967; Giomo, 1993; Cohen, 1990 (noncontinuous bipolar); Schubert, 1999b (continuous bipolar).

Comparing the evaluation of the measures presented above with the criteria for a good measure of musical emotions (as described at the beginning of this chapter), the rating scales appear to provide a better balance between all the requirements (as also supported by Schubert (1999a)). They are the better method for obtaining a continuous quantity, as they overcome the problem of semantic density from other measures, and they can also be attached to meaningful qualities of the emotion. Still two main aspects are undesirable: the practical limitations on the number of dimensions measured simultaneously (especially for continuous quantities) and the restrictions caused by them on the "quality" of the emotion description and its meaningfulness. Nevertheless, repeatedly, different empirical studies have provided consistent evidence of a reduced number of dimensions (or semantic scales) suitable to describe the subjective experience of emotion.

---

the list. Usually the procedure continues until all were selected. With ranking measures, instead of selecting words from the list, these words are ranked accordingly to their similarity to describe the emotional experience or value. They often appear mixed with other kinds of measures, and are rarely used alone.

Dimensional theories

According to Wundt (1896), differences in the affective meaning among stimuli can succinctly be described by three pervasive dimensions (of human judgement): pleasure ("lust"), tension ("spannung") and inhibition ("beruhigung"). This model has received empirical support from several studies, which have shown that a large spectrum of perceptual and symbolic stimuli can be represented using these dimensions (see Bradley & Lang, 1994). They can be represented in a three-dimensional space, with each dimension corresponding to a continuous bipolar rating scale: pleasantness-unpleasantness, rest-activation and tension-relaxation.

Other studies have provided evidence that the use of only two dimensions is a good framework to represent affective responses to linguistic (Russell, 1980), pictorial (Bradley & Lang, 1994) and musical stimuli (Thayer, 1986). These dimensions are labelled as arousal and valence. Arousal corresponds to a subjective state of feeling activated or deactivated. Valence stands for a subjective feeling of pleasantness or unpleasantness (hedonic value) (Russell, 1989).

By achieving a meaningful representation of emotion with only two scales, the main drawback of bipolar ratings scales is reduced. They are also suitable for continuous measurement frameworks, supporting the analysis of the time course of emotion in more detail. They also permit the representation of a very wide range of emotional states, since they describe a continuous space where each position has no label attached (the only restriction is the meaning of each scale). This is particularly important in the context of musical emotions. Its simplicity in terms of psychological experiments and good reliability (Scherer, 2004) have consistently promoted its use in emotion research.

Another important aspect is to locate the participant's responses within the mechanisms of emotion. Participants are asked to either focus on the

experienced or the perceived emotions (Gabrielsson, 2002). In any case, their own feelings of emotion (emotion "felt") or the emotion known to be represented by the music, are the reported ratings. In this way the framework focuses on the subjective feelings of emotion, without considering, at least explicitly, other components of the emotional experience. Some studies have shown that these dimensions may also relate to physiological activation. Lang and other colleagues (Greenwald, Cook, & Lang, 1989; Lang, Bradley, & Cuthbert, 1998) have recently compiled a large database indicating that cardiac and electrodermal responses, and facial displays of emotion, show systematic patterns in affect as indexed by the dimensions of arousal and valence ("pleasure"). Feldman (1995) also suggests that the conscious affective experience may be associated with a tendency to attend to the internal sensations associated with an affective experience.

Schubert (1999a) has applied this concept to music creating the EmotionSpace Lab experimental software. While listening to music, participants were asked to continuously rate the emotion "thought to be expressed" by music. Each rating would correspond to a point in the Arousal/Valence dimensional space. This approach overcomes some of the drawbacks related to other techniques which do not take into consideration changes in emotion during music (Sloboda, 1991), and it also supports the study of the interaction between quantifiable and meaningful time varying features in music (psychoacoustic dimensions) and emotion (Arousal and Valence). Several other authors (e.g. Grewe, Nagel, Kopiez, & Altenmuller, 2005; Korhonen, 2004a) have also used this model to find interactions between ratings of Valence and Arousal and the acoustic properties of different pieces.

2.4.2 Physiological measurements

When a stimulus or event of value is perceived, the body is prepared for action, partially through increased physiological arousal. Some of these responses involve the activation of the sympathetic nervous system, but also the release of arousal hormones. Increased arousal is associated with increased heart rate, increased body temperature, increased respiration rate (increased oxygen consumption), decreased skin temperature and many other physiological changes. From a neurological perspective, increased arousal is associated with the release of noradrenaline and cortisol, affecting for instance the amygdala. As a reference, the amygdala, deep in the limbic system, can influence cortical areas via feedback from proprioceptive, visceral or hormonal signals, via projections to various networks. This means that hormones secreted in the body affect bodily processes (e.g. cardiovascular, muscular and immune systems) and the brain as well.

As reviewed by Dainow (1977), right from early studies that investigate physical and motoric responses to music to the most recent ones, one of the most common measures used is heart rate variations. Ever since Grétry's first recordings in the 18*th* century (see Dainow, 1977), most researchers have used this variable as a measure of physical responses to music. Throughout the years, the connection between the rhythms of music and body received substantial support and provided evidence from different areas of research (e.g. musicology, experimental psychology and even neuroscience). For instance, the average tempo in music is considered to be about 72-80 beats per minute, a value very similar to the average heart rate of a (healthy) person. The similarities of such values with others such as the timings of perception or even with the general rate of brain activity (see McLaughlin, 1970), has led researchers to explain this connection. Results often pointed in different directions, and strong beliefs such as the hypothesis that the heart rate follows the music tempo have never

been proved. Although with less preponderance, the respiration rate, respiration amplitude (still the body rhythms), and measurements of skin conductance, are also commonly used. A good review of the relationships between internal rhythms and external drivers was written by Byers (1976).

The lack of experimental techniques and statistical analysis procedures was a significant limitation in physiological research before the 20th century, which could partially explain the diversity of results found in those studies. Current technology permits the measurement of for a wider range of parameters, but the general knowledge about the physiological effects of music is still to be unveiled.

More recently, several researchers have used physiological measurements to quantify the relationship between affective states and bodily feelings (e.g. Khalfa et al., 2002; Rickard, 2004; Krumhansl, 1997; Witvliet & Vrana, 1996; Harrer & Harrer, 1977). Krumhansl (1997) measured a large number of variables related to the cardiac, vascular, electrodermal, and respiratory functions). Participants reported (on continuous unipolar scales) the level of a few emotion quality ratings (happiness, sadness, fear, tension). The analysis of the relationships between both measures of emotion has shown that *sad* excerpts produce the largest changes in heart rate, blood pressure, skin conductance and temperature. The *fear* excerpts produced the largest changes in blood transit time and amplitude, while the happy excerpts induced the largest changes in the respiration variables. The strongest correlations found corresponded to the skin conductance level (SCL).

In another study, Khalfa et al. (2002) found that the skin conductance level increases for arousing or emotionally powerful music, supporting the idea that physiological arousal is a component of musical emotions. Due to its direct connection with the sympathetic nervous system, the fluctuations in the eccrine sweat gland activity convey an important source of measurements about its activity. In this context, some authors claim that the degree of bodily

arousal distinguishes intense emotions from less intense emotions (the "arousal hypothesis").

Iwanaga and Tsukamoto (1997) examined in more detail the excitative-sedative effect of music through spectral analysis of the heart rate. The subjective feeling of musical activity was measured using an adjective checklist, and compared with heart-rate variabilities divided into two groups: low frequency - mainly affected by the sympathetic nervous system - and high frequency - mainly affected by the parasympathetic nervous system.

It was observed that the excitative-sedative effect of music were only differentiated in indices related to high frequencies, suggesting that the musical effect may be observed in measures of the parasympathetic nervous system but not in the sympathetic nervous system[7].

Peripheral feedback in music

Dibben (2004) has applied the peripheral feedback ideas to music. In a double experiment, she investigated the role of physiological arousal in determining the intensity and hedonic value of the emotion experienced while listening to music (the peripheral feedback hypothesis).

In one experiment Dibben had two groups of participants with different levels of induced physiological arousal (participants who exercised before reporting the feelings of emotion, and another with participants that just relaxed). Participants were asked to report the emotion "felt" while listening to the music, as well as the emotion "thought to be expressed" by the music. Dibben found that the groups of participants with induced physiological arousal reported more intense emotions "felt" more than those who didn't, whereas increased physiological arousal intensified the dominant valence of the emotion "felt". No effect was

[7]These two systems correspond to different divisions of the autonomic nervous system (a subdivision of the peripheral nervous system). Both are related to the control of smooth muscle contraction, the regulation of the cardiac muscle, and the stimulation or inhibition of glandular secretion.

found for the report of emotion "thought to be expressed" by the music.

The second experiment was carried out to study the effect of different types of induced physiological arousal on emotion, with separate groups rating emotion "felt" and emotion "thought to be expressed" by the music. Dibben also looked at the origin of the effect, whether it was due to peripheral feedback or due to mood changes associated with the experiment. This experiment has shown a stronger effect of physiological arousal for emotions "felt" and "thought to be expressed", for music positive in valence. No differences were found in the participants mood state as a consequence of induced physiological arousal. Dibben also found significant relationships between reports of emotion, self reported and physiological arousal.

Dibben's work provides 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.

Chapter 3

Computational modelling in music and emotions

A computational model is a computer program that attempts to simulate an abstract model of a particular system. Often, these models are mathematical formalisation of the system (or groups of systems) they represent. The main goal for creating these models is to achieve a better understanding of the system they aim to represent, often to interpret its underlying mechanisms.

The initial stages of the modelling process involve the development of explicit mathematical descriptions of the system in study, such as, a mathematical model of the ear, or the way the brain "behaves" when we perform an action, among many others. The next step is the transformation of this formalisation into a computational model. That is achieved by creating computer programs of the mathematical descriptions.

A computational model, being an abstract representation of the system under study, can be used as a platform to investigate some aspects of its nature. Such studies, or "in silico experiments", are commonly referred to as simulations. The intended outcome of the simulations carried out with a model can also be compared to empirical data. That means that, if the model is valid, it can also

be used to test the validity of the theory that it represents or even to produce new hypotheses, which in turn can be used for the development of new empirical experiments.

In the following section I will describe some studies that have made use of computational models to study different aspects of affective responses or emotion related phenomena. The emphasis will be on those works that investigate emotional responses to music, but I will also describe some works that have applied mathematical and modelling techniques to generate or reshape certain musical and sound features. These also emphasise the importance of the organisation of separate musical elements in the process of composition.

3.1 The use of computational models in music

Computational models have been used in many different areas of music research. Some of these include techniques for sound design and composition (Mandelis & Husbands, 2003; Xenakis, 1971), models of creativity (Manzolli, Moroni, Zuben, & Gudwin, 1999; Cope, 1991), computational musicology (Gimenes & Miranda, 2008; Coutinho, Gimenes, Martins, & Miranda, 2005), music information retrieval (e.g Goto, 2004; Feng, Zhuang, & Pan, 2003), auditory modelling (e.g. Rosenthal & Okuno, 1998; Lyon, 1984), models of tonality (Cambouropoulos, 2003; Bharucha, 2002; Tillmann, Bharucha, & Bigand, 2000), among many others. In this chapter, I will first describe some of these models. Then, I will focus on the relationships between sound features and emotional states from a computational model perspective (the scope of this thesis). As surveyed in the previous chapters, both music and emotion are characterised by important temporal features and relationships between those features, and so their study requires models that are capable of representing these relationships. The use of spatiotemporal connectionist models will be proposed as a good paradigm

for their study. The methodology of artificial neural networks and specifically spatiotemporal models will be described in detail in this chapter.

3.1.1 Creativity and improvisation

The production of sound faced a revolution in the middle of the 20^{th} century with the appearance of the digital computer (Mathews, 1963). Computers were given instructions to synthesise new sounds algorithmically. Synthesisers (or *software synthesisers*) soon became organised as a network of functional elements (signal generators and processors) implemented in software.

Composers have used a number of mathematical models such as combinatorial systems, grammars, probabilities and fractals (Dodge & Jerse, 1985; Cope, 1991; Worral, 2001) to compose music that does not imitate well known styles. Some of these composers created very interesting pieces of new music with these models and opened innovative ground in compositional practices, such as the techniques created by Xenakis (1971).

The use of emergent and generative methods (e.g evolutionary algorithms (EA)) is another trend that is becoming very popular for its potential to generate new music of relatively good quality. A great number of experimental systems have been used to compose new music using EA: Cellular Automata Music (Millen, 1990), CA Music Workstation (Hunt, Kirk, & Orton, 1991), CAMUS (E. R. Miranda, 1993), MOE (Degazio, 1999), GenDash (Waschka II, 1999), CAMUS 3D (McAlpine, Miranda, & Hogar, 1999), Living Melodies (Dahlstedt & Nordhal, 2001) and Genophone (Mandelis, 2001), to cite but a few.

For example, *CAMUS* (E. R. Miranda, 1993) takes the emergent behaviour of Cellular Automata (CA) to generate musical compositions. This system, however, goes beyond the standard use of CA in music in the sense that it uses a two-dimensional Cartesian representation of musical forms. In this representation the coordinates of a cell in the CA space correspond to the distances between the

notes of a set of three musical notes.

As for EA-based generative music systems, they generally follow the standard genetic algorithm procedures for evolving musical materials such as melodies, rhythms, chords, and so on. One example of such a system is *Vox Populi* (Manzolli et al., 1999), using computational procedures to evolve populations of chords of four notes, through the operations of crossover and mutation.

EA have also been used in systems that permit for interaction in realtime, i.e., while the composition is being generated. In fact, most EA-based systems have the advantage of letting the user control EA operators and fitness values while the system is running. For example, Impett (2001) proposed an interesting swarm-like approach to interactive generative musical composition. Musical composition is modelled here as an agent system consisting of interacting embodied behaviours. These behaviours can be physical or virtual and they can be emergent or preset. All behaviours coexist and interact in the same world, and are adaptive to the changing environment to which they belong. Such behaviours are autonomous, and prone to aggregation and generation of dynamic hierarchic structures.

3.1.2 Computational Musicology

Computational musicology is broadly defined as the study of Music by means of computer modelling and simulation. Artificial Life models and EA are particularly suitable to study the origins and evolution of music. This is an innovative approach to a puzzling old problem: if in Biology the fossils can be studied to understand the past and evolution of species, these "fossils" do not exist in Music; musical notation is a relatively recent phenomenon and is most prominent only in the Western world. "*In silico*" simulation are suggested to be useful to develop and demonstrate specific musical theories.

Todd and Werner (1999) proposed a system for studying the evolution of musical tunes in a community of virtual composers and critics. Inspired by the

notion that some species of birds use tunes to attract a partner for mating, the model employs mating selective pressure to foster the evolution of fit composers of courting tunes. The model can coevolve male composers who play tunes (i.e., sequences of notes) along with female critics who judge those songs and decide with whom to mate in order to produce the next generation of composers and critics. This model is remarkable in the sense that it demonstrates how a Darwinian model with a pressure for survival mechanism can sustain the evolution of coherent repertoires of melodies in a community of software agents.

Miranda (E. Miranda, 2004; E. Miranda, Kirby, & Todd, 2003) proposed a mimetic model to demonstrate that a small community of interactive distributed agents furnished with appropriate motor, auditory and cognitive skills can evolve from scratch a shared repertoire of melodies (or tunes) after a period of spontaneous creation, adjustment and memory reinforcement. One interesting aspect of this model is the fact that it allows us to track the development of the repertoire of each agent of the community. Metaphorically, one could say that such models enable us to trace the musical development (or "education") of an agent as it gets older.

3.1.3 Expression in music performance

One example of how performance expressivity might be modelled was proposed by Juslin, Friberg, and Bresin (2002). This system decomposes patterns of expression into different subcomponents. The authors propose that expression derives from four primary components: generative rules, which mark the structure in a musical manner; emotional expression, which serves to convey particular moods; random fluctuations, which reflect human limitations in timing precision; and motion principles, which postulate that tempo changes should follow patterns of human movement.

In the context of Western tonal music, some of the music expressions are

delivered in a performance by subtle deviations from the notated musical score. Focusing on this aspect, several studies of expressive music performance have aimed to establish why, where and how these deviations take place in a piece of music. Some authors (e.g. Bresin, 1998; N. Todd, 1992) have explored computational models in order to "connect" the properties of a musical score and performance context to the physical parameters of a performance, such as timing, loudness, tempo, articulation. Nonetheless, several other strategies have also been employed (e.g., analysis by measurement, analysis by synthesis, machine learning and so on) in order to capture common performance principles. Refer to (Widmer & Goebl, 2004) for further details and related references.

3.1.4 Musical instruments and emotion

Suzuki and Hashimoto (1997) modelled human evaluations of tones played on different instruments. An "emotional sound space" was constructed based on subjective comparisons between sounds of different musical instruments. In this work the "emotional appraisals" are not a direct measure. Instead, they emerge from the similarity between pairs of instruments. The authors' intention was to map the timbre of an instrument to a two (or three) dimensional emotion space using similarity measures. The researchers recorded the sound of a single tone from twenty-two different instruments (one and a half seconds each) and computed the power spectrum of the audio data as a function of time, using a multidimensional vector (128 dimensions). Using Principal Component Analysis (PCA) they reduced the dimensionally of this vector.

Each participant in the experiment listened to all the possible pairs of ten of the instruments and rated the similarity of their subjective feelings of emotion to both instruments (using a seven point Likert scale score from "similar" (1) to "not similar" (7)). Then the authors created a three layer perceptron neural network to estimate the placement of the instruments in a multidimensional space (in

such a way that the similarity measures between pairs of instruments correspond to the Euclidean distance between them). The input to the neural network is a multidimensional vector representing the audio data of ten instruments, and the output is the location in the "emotional space". The network was trained to reproduce the participants ratings of the similarity between two instruments, and so to estimate the Euclidean distance between the instruments in the emotion space. The remaining twelve instruments were used to test the generalisation capabilities of the model.

Although Suzuki and Hashimoto (1997) were able to generalise their "emotional sound space" to all the instruments, the emotional dimensions used have no direct interpretation or meaning. Moreover the model has only been tested with single tones, which leaves aside more realistic scenarios (multiple notes from multiple instruments). Although not very informative in terms of the emotional values of timbre, this work has provided a very interesting framework to represent timbre based on the subjective impression of listeners.

3.1.5 Classification of music selections

Li and Ogihara (2003) modelled emotional appraisals of thirty second music excerpts for four genres of music. The emotional appraisals consisted of a multilabel classification problem. The "emotion detection" process was divided into two phases: feature extraction and multilabel classification. In the first step, thirty features were extracted from the music excerpts using Marsyas (Tzanetakis & Cook, 2000). The features selected intended to represent the musical properties of timbre, rhythm and pitch. In the second step, all the excerpts were labelled by a subject (a 39 year old, male), using ten adjective groups proposed by Farnsworth (1958). The subject was instructed to select all adjective groups that approximated the subjective feeling of each excerpt (there was no limit to the number of groups chosen, and the subject could also

suggest new adjective groups if necessary). The classifiers consisted of thirteen emotions and six emotion "supergroups". Support vector machines were used as the classifiers, and the inputs were the music features vector. While half of the excerpts were used to train the model, the remaining data was used to evaluate the generalisation performance.

Although the modelling framework was able to generalise emotional appraisals of musical selections from a variety of genres of music, this study cannot be used to generalise emotional appraisals of music for a population of listeners, because only one listener was used in the experiment. Moreover, the reported performance of this model was poor, and the researchers assumed that the emotional appraisals were fairly constant over each excerpt.

Feng et al. (2003) implemented a music retrieval system based on mood detection. Mood detection was implemented by analysing tempo and articulation, which was assumed to be constant throughout each piece. The authors derived four categories of mood from those excerpts: happiness, anger, sadness and fear. Then a feedforward neural network classifier was trained to detect mood. During the network training procedure, the inputs consisted of the features derived from the music signal and the outputs corresponded to the mood scores (obtained from a musicians estimation) of each music piece. After training, the output scores represent the estimations of the mood evoked. The final system was then tested in an experimental context, where users can select pieces associated with specific moods. The precision of the model in selecting the desired mood is 67%, nevertheless most parts of the tests were performed on the training data (330 pieces). Because only 23 pieces were used as novel data it is not clear how well this model performs. Like the previous model, this system also assumes that the music features are constant throughout the music excerpts, which is not plausible.

3.1.6 Controlling music emotionality

Livingstone (2007) presented an affective computing architecture to modify pre-existent music in order to achieve a predictable affective value. The author aims to create a model that provides reliable control of both perceived and induced musical emotions. He uses a rule-based system to modify a subset of musical features at two levels: score and performance. The model adapts the emotionality of the music by modifying it in real-time, aiming to allow the listener to reach a desired emotional state. The author does not provide an assessment of their work.

A similar approach is taken by Oliveira and Cardoso (2008). They propose a conceptual system that can control the affective content of precomposed music (represented at a symbolic level). The aim is to adapt the piece to an intended emotional description, using a precompiled knowledge base with weighted mappings between continuous affective dimensions (valence and arousal) and musical features (rhythm, melody, etc.). The system consists of the segmentation, analysis, selection, transformation, sequencing, remixing and synthesis of precomposed symbolic music, and its assessment consists of playing produced music and identifying emotions of the listener.

The suggested implementation starts with the choice of an emotional description (specified by the listener). Then, the rules associated with the emotional description are selected from the knowledge base. These are used to select a piece from a music database of MIDI (White, 2000) files, according to similarity metrics between music features. Then, the selected music can be transformed, sequenced, remixed and synthesised using specific algorithms. To calibrate the system the emotions identified are compared to the intended emotions in an experimental context where listeners rate the affective value of the selected music. These comparisons are used to refine the mappings in the knowledge base. Because this model has not yet been implemented it could not

be evaluated.

3.1.7 Emotional responses and electrical activity in the brain

When the power spectrum of sequences of musical notes is inversely proportional to the frequency on a log-log plot, it is called 1/f music. According to Voss and Clarke (1978), most listeners agree that 1/f music is much more pleasing than white (1/f0) or brown (1/f2) music, which sound either too "random" or too "correlated", respectively. The studies suggests that musical pleasure depends not so much on the absolute value of perceptual elements (e.g. the pitch, the tone duration or the loudness) but rather on how it changes as a function of time.

Based on these premises, Jeong, Joung, and Kim (1998) studied the way in which the emotional response to music is reflected in the electrical activities of the brain. Jeong et al. (1998) used nonlinear methods to investigate the chaotic dynamics of electroencephalograms (EEGs) elicited by computer generated $1/f$, white ($1/f^0$), and brown ($1/f^2$) music. In this analysis, the authors used the correlation dimension and the largest Lyapunov exponent as measures of complexity and chaos. A new method was also developed for calculating the nonlinear invariant measures from limited noisy data.

Their findings show that at the right temporal lobe, 1/f music elicited lower values of both measures (correlation dimension and largest Lyapunov exponent) than $1/f^0$ (white) or $1/f^1$ (brown) music. It was also observed that "brains which feel more pleased" show decreased chaotic electrophysiological behaviour, with rhythm variations having a greater contribution to pleasurable responses to music than the melody ones. Overall, Jeong et al. (1998) provide new insights on the dynamical mechanism of music perception and contribute to the modelling of emotion using nonlinear techniques.

3.1.8 Time series analysis of music and emotions

Only a few studies have focused on modelling emotional responses to music focusing on the perceived sound patterns. In these studies the music is encoded into time varying quantities describing the psychoacoustic experience (sound features). The sound features correspond to perceptually separable elements (or groups of elements), universal to musical sound, that when combined form the "musical object". They are used to create models that are able to describe at some level the affective experience with music. As time varying variables they give us an idea of the variation occurring in the different perceptual dimensions throughout the music, and view of emotion with an appropriate time scale.

Within this framework two models of continuous measurements of emotion (using psychoacoustic variables and time series analysis) were proposed by Schubert (2004) and Korhonen (2004a). Both authors attempted to model the interaction between music psychoacoustics and emotional ratings, focusing on the continuous response methodology which takes into account the variations in emotion as a piece of music unfolds in time. Schubert proposes a methodology based on the combinations of time series analysis techniques to analyse the data and to model such processes. Korhonen proposes the use of System Identification (Ljung, 1986).

Schubert applied an ordinary least squares stepwise linear regression model (using sound features as predictors of the emotional response) and a first order autoregressive model (to account for the autocorrelated residuals and providing the model with a kind of "memory" of past events) to his experimental data. For each piece, Schubert created a set of models of emotional ratings (arousal and valence) and selected musical features (melodic pitch, tempo, loudness, frequency spectrum centroid and texture). Each sound feature was also lagged (delayed from the original variable) by 1 to 4 s. Schubert assumption was that the emotional response will occur close to or a short time after the causal musical

event.

The modelling technique used by Schubert's suffers from a number of drawbacks. First, it assumes that the relationships between music and emotional ratings are linear. This is a very optimistic view taking into account the nature of the neural processes involved in sound perception. Second, the relationships between sound features and emotional response are considered to be mutually independent. This factor is particularly restrictive, since it discards altogether the interactions between sound features. This is an oversimplification of the relationships between sound features and emotional response, and an acknowledged limitation in music and emotion studies (Gabrielsson & Lindström, 2001). The interactions between variables are a prominent factor in music. As also concluded by Schubert, more sophisticated models that can account for more detailed descriptions of the relationships between the dynamics qualities of music structure and emotions are needed to better understand the nature of this process. Another limitation of Schubert's work is lack of prediction of the emotional responses to novel music, since he created separate models for each piece and each affective dimension. Moreover, the relationships found between sound features and affective response for the different piece were piece specific (sometimes even contradictory), not assuring the model validation of model.

In the second study, Korhonen (2004a) extended these experiments and address some of Schubert's limitations. The sound features space and the musical repertoire were increased, in order to incorporate more music and psychoacoustic (sound) features. The modelling techniques include all the music variables in a single model and the generalisation to new music is also tested. Nevertheless, the interactions between sound features and the contribution of nonlinear models are issues still not addressed.

System identification describes a set of mathematical tools and algorithms that create models from measured data. Typically, these models, are either

based on predefined (although adjustable) model structures (e.g. state-space models), or with no prior model defined (e.g. neural network). Korhonen considered state-space models, ARX and MISO ARX models (the last using delays estimated automatically from the step response - see Korhonen, 2004a for further details). The best model (the one with better generalisation performance - the average of six models using all six songs as estimation data) reported explains 7.8% of valence and 75.1% of arousal responses. The performance for valence is very poor, even though only one piece at a time was used for generalisation. Comparing with Schubert's results, Korhonens models showed worst performance for some pieces and better for other (particularly worst for valence). It also used more sound features including tempo, loudness, texture, mean pitch, harmony (fifteen in total).

In a followup study, Korhonen, Clausi, and Jernigan (2004) also assessed the contribution of feedforward neural networks with input delay elements, and state-space models. Each model was again evaluated by the average performance of six models testing the generalisation for each piece (using five of the pieces to estimate the parameters of the model). State-space models were again unsuccessful at modelling valence (suggesting that linear models may not be appropriate to estimate valence). The neural network model used improved the model performance only for valence, explaining 44% of the response. Again the generalisation is only for one piece at a time.

These two studies are closely related to this research. They constitute a starting point for my modelling and computational investigations.

3.2 Connectionism and Artificial Neural Networks

Connectionist models, also known as Artificial Neural Networks (ANN), are adaptive (most often nonlinear) systems that learn to perform a function (an

input/output map) from data. Adaptive means that the system parameters are changed during operation. This modelling paradigm is based upon a network of interconnected simple processing units that attempt to model information processing in the way it actually takes place in the brain (Rumelhart, Hintont, & Williams, 1986). This modelling principle was developed upon the discovery a system of neural connections appeared to be distributed not only in serial pathways, but also in parallel arrays, suggesting that different types of neural processing are distributed throughout complex systems of neural networks. Connectionist models have the particularity of processing information differently from traditional computers: the computation occurs in parallel, rather than in a serial fashion as in traditional computer architectures. Such a process is often called parallel distributed processing (PDP), emphasising the nature of the computation. This type of information processing has eight basic components(Rumelhart et al., 1986):

- A set of processing units;
- A state of activation;
- An output function (for each unit);
- A pattern of connectivity among units;
- A propagation rule for spreading the activation patterns through the network;
- An activation rule for combining the inputs which have an effect on specific unit with the current state of that same unit (generating the next state for the unit);
- A learning rule to modify the patterns of connectivity (weights);
- An environment within which the system must operate.

This type of model paradigm is very flexible in terms of application because it offers a highly personalised definition of the model characteristics. The processing units (called artificial neurons) are arranged in various possible topologies, which define the way the artificial neurons interact and the flow of information within the model. Their creation involves a training phase and a testing phase. Learning algorithms (supervised or unsupervised) define the way the model behaves during training when tuning its parameters for a certain task. They consist of systematic step by step procedures to optimise the performance of the model in relation to a predefined criterion. The test phase serves to test the model with novel data, either for predictions or validation of the model. A typical example of neural network training is categorisation: given a set of training stimuli, the model is asked to separate them into a predetermined set of categories. An interesting phenomenon arises when we present the system with new stimuli never seen previously. These new inputs, after a successful learning process, should ideally be categorised within the learned categories space, reflecting the underlying grammar of the process being modelled. This process is called generalisation and it allows ANN's to be used in unknown environments. The following sections give an overview of the main components of a connectionist model.

3.2.1 Processing units

Processing units are the basic blocks of ANNs. They are responsible for the information processing which goes on within the network. Each unit receives input from other units or external sources using this information to compute an output signal that is propagated to other units. Within neural systems, there are three main types of processing units: the input units, which receive data from external sources (outside the neural network), the output units, which connect to the external world by sending the data out of the neural network, and the hidden units,

whose input and output signals remain within the neural network. The system is inherently parallel because different units can process their computations at the same time.

The processing units are based upon a set of subcomponents: an input function, an activation (or transfer) function and a state of activation (equivalent to the output value of the unit). The input function determines the signal that the processing unit receives from all its sources. Most often this function it is just the weighted (w_{jk}) sum of input signals that connect to a specific unit (k) plus a bias value (θ_k - which behaves as an *offset*). The contribution of an input with positive weight ($w_{jk} > 0$) is also known as an excitatory connection, while a negative one is called inhibitory ($w_{jk} < 0$).

$$s_k(t) = \sum_j w_{jk}(t) y_j(t) + \theta_k(t) \tag{3.1}$$

The output of this computation ($s_k(t)$) is then passed through the unit activation function (F_k), which in turn outputs the activation values of the unit. The rule used is typically some sort of threshold function, such as hard limiting threshold functions (e.g. a sgn function), linear or semilinear functions (e.g. linear function limited in the edges), or a smoothly limiting threshold (e.g. a sigmoid function) - see Figure 3.1.

$$y_k(t+1) = F_k(s_k(t)) = F_k\Big(\sum_j w_{jk}(t) y_j(t) + \theta_k(t)\Big) \tag{3.2}$$

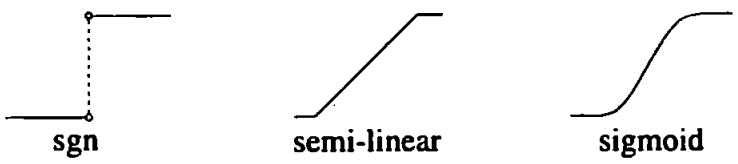


Figure 3.1: Typical activation functions used in the units (artificial neurons) of artificial neural networks.

3.2.2 Topologies

ANN topologies define the pattern of connections between the units and the propagation of data within the network. There are two general types of topologies: feedforward and recurrent. In the first type, the data only flows from the input to output units, but the data processing can be extended over multiple layers (groups of processing units). There are no connections from output to input and between units in the same layer or previous layer. Instead, the second topology, can contain feedforward connections as well as the other two types, which are also called feedback connections. Examples of feedforward neural networks are the (singlelayer or multilayer) Perceptron (Rosenblatt, 1958) and the Adaline (Widrow & Hoff, 1960). Examples of recurrent topologies are the ones presented by Kohonen (1977), Hopfield (1982) , Jordan (1990) and Elman (1990) .

3.2.3 Training and learning in neural networks

The strengths of the connections of a neural networks have to be configured in such a way that, within a specific application, a set of inputs produces the desired set of outputs. One way is to set the weights explicitly using previous knowledge or the system. Another, more common, is to "train" the neural network by letting it change its weights according to some learning rule.

There are two main classes of learning rules that can be applied to neural networks: supervised (or associative) and unsupervised (or selforganisation). When using supervised learning, the network is trained by providing it with a set of inputs and the set of desired outputs that match those inputs. The input/output pairs can be provided by the user or another program, or even by the system in which the network is embedded. With unsupervised learning instead, the outputs are trained to respond to patterns within the inputs: the network, associated with a learning algorithm, is supposed to "discover" statistically salient features on the

input data, with no "a priori" knowledge of how to classify those patterns.

There are many algorithms available for training artificial neural networks (some of them are specific for some topologies): hebbian learning, principal component analysis, evolutionary methods (e.g. genetic algorithms), simulated annealing, reinforcement learning, expectation maximisation, competitive learning, among many others (and variants of these). Nevertheless, the most popular in multilayered artificial neural networks (and relevant for this thesis) is the backpropagation (BP) algorithm. Using a BP algorithm, the output values are compared with the desired (target) to calculate the value of some predefined error function (e.g. the delta rule). This error is then fed back through the network (backpropagation). The backpropagated error values are then used to adjust the weights of each connection weight in order to reduce the value of the error function by some amount. This process is repeated for a sufficiently large number of training cycles, and the network is expected to converge to some (desired) state where the difference between the outputs and the desired values is minimised. In a successful training process the network "learns" a certain target function. The adjustment of the connections weight is achieved through a method for nonlinear optimisation: the gradient descent (because of that a backpropagation algorithm can only be applied on networks with differentiable activation functions). This method takes the derivative of the error function with respect to the network weights. Those weights are then changed such that the error decreases.

3.3 Spatiotemporal connectionist models

The intention of creating a computational model is not only to represent a desired system, but also to achieve a more refined understanding of the underlying mechanisms. In this thesis the hypothesis for the model is that sound features

convey information along two affective dimensions: arousal and valence. In that way the knowledge to be extracted from the model should reveal some information about its underlying mechanisms. Then it would be possible to generate new hypotheses and make predictions for new data.

In order to model continuous measurements of emotional appraisal I will consider the contribution of spatiotemporal models, i.e. approaches where the model at the same time includes a temporal dimension (e.g. the dynamics of musical sequences and continuous emotional ratings) and a spatial component (e.g. the parallel contribution of various music and psychoacoustic factors). A spatiotemporal connectionist model (Kremer, 2001) can be defined as "a parallel distributed information processing structure that is capable of dealing with input data presented across time as well as space" (Kremer, 2001, pp. 2).

Both conventional and spatiotemporal connectionist networks (STCN) are equipped with memory in the form of connection weights. These are represented as one or more matrices, depending on the number of layers of connections in the network. These weights are typically updated after each training step and constitute a memory of all previous training. This memory extends back past the current input pattern to all the previous input patterns. In this sense it usually referred to as long-term memory. Once a connectionist network has been successfully trained, it remains fixed during the operation of the network.

An additional characteristic of STCNs is that they also include a form of short-term memory. It is this memory that allows STCNs to deal with input and output patterns that vary across time as well as space. While conventional connectionist networks compute the activation values of all nodes at time t based only on the input at time step, in STCNs the activations of some of these nodes is computed based on previous activations, serving as a short-term memory. Unlike the weights (long-term) memory, which as explained remain static once the training period is completed, the short-term memory is continually recomputed with each

new input vector in both training and operation.

As mentioned earlier, recurrent neural networks, involve various forms of recurrence (feedback connections). Through these, some of the information at each time step is kept as part of the input to the following computational cycle. By allowing feedback connections, the network topology becomes more flexible. It is possible to connect any unit to any other, including to itself. At each time step the activations are propagated forward through one layer of connections only. Then the memory units are updated with their new states. This information will continue to flow around the units, even in the absence of any new input whatsoever. These models 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. Some of the applications of these models are the learning of formal grammars (e.g. Lawrence, Giles, & Fong, 2000; Elman, 1990), spoken word recognition (McClelland & Elman, 1986), written word recognition (Rumelhart & McClelland, 1986), speech production (Dell, 2002), and music composition (e.g. Mozer, 1999).

There are various proposals and architectures for time-based neural networks (see Kremer, 2001 for a review) making use of recurrent connections in different contexts. Examples of these models are the Jordan Network (Jordan, 1990) and the Elman Neural Network (ENN) (Elman, 1990). Jordan and Elman neural networks are extensions of the multilayer perceptron, with additional context units which "remember" past activity. These units are required when learning patterns over time (i.e., when the past computations of the network influence the present processing). The approach taken by Jordan (1990) involves treating the network as a simple dynamic system in which previous states are made available to the system as an additional input. During training, the network state is a function of the input of the current time step, plus the state of the output units of the previous time step.

By contrast, in the Elman network, the network's state depends on the current inputs, plus the model's internal state (the hidden units activations) of the previous cycle. This is achieved through an additional set of units (memory or context units), which provide (limited) recurrence. These units are activated on a one for one basis by the hidden units, copying their values: at each time step the hidden units activation are copied into the context units. In the following cycle, the context values are combined with the new inputs to activate the hidden units. The hidden units map the new inputs and prior states to the output. Because the hidden units are not trained to respond with specific activation values, they can develop representations in the course of learning. These encode the temporal structure of the data flow in the system: they become a "task specific" memory (Elman, 1990). Because they themselves constitute the prior state, they must develop representations that facilitate this input/output mapping.

3.3.1 Motivation for the use of Elman Neural Networks

The dynamic qualities of music are perhaps its most important, due to the fact that music is characterised by constant changes over time. The need to study in more detail the dynamic aspects of affective responses to music led to a focus on continuous measurements frameworks to investigate emotion, shifting the attention from mood studies to studies on emotion. The use of continuous measurements is also motivated by the idea that musical emotions may exhibit time locking to variations in psychological and physiological processes, consistent with a number of studies that show temporal variations in affective responses (e.g. Goldstein, 1980; Nielsen, 1987; Krumhansl, 1997; Schubert, 1999a; Korhonen, 2004a). Because the static attributes of music are only partially responsible or indicative of emotional response to music, which can be intense and momentary (e.g. Dowling and Harwood (1986)), the study of its dynamics in the context of time series analysis needs to be explored.

For this study the Elman network (Elman, 1990) was chosen. The fundamental additional aspect of an Elman Neural Network (ENN) (when compared to the traditional feedforward model) is the use of recurrent connections that endow the network with a dynamic memory. This way the network can also detect temporal patterns in the data, at different time lags. The internal representations of an ENN encode not only the prior event but also relevant aspects of the representation that was constructed in predicting the prior event from its predecessor (thats the effect of having learned weights from the memory to the hidden layer). The basic functional assumption is that the next element in a time series sequence can be predicted by accessing a compressed representation of previous hidden states of the network and the current inputs. 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. During learning, the hidden units must accomplish an input-output mapping and simultaneously develop representations that systematic encodings of the temporal properties of the sequential input at different levels (Elman, 1990). In this way, the internal representations that drive the outputs are sensitive to the temporal context of the task (even though the effect of time is implicit). The recursive nature of these representations (acting as an input at each time step) endows the network with the capability of detecting time relationships of sequences of features, or combinations of features, at different time lags (Elman, 1991). This is an important feature of this network because the lag between music and affective events has been consistently shown to vary in the order of 1 to 5s (Schubert, 2004; Krumhansl, 1996; Sloboda & Lehmann, 2001). Another important aspect is that ENNs have very good generalisation capabilities. This technique has been extensively applied in areas such as language (e.g. Elman, 1990) and financial forecasting systems (e.g. Giles, Lawrence, & Tsoi, 2001), among others.

Chapter 4

A spatiotemporal neural network model of musical emotions: Modelling subjective feelings

This simulation will consist of the training of an ENN to learn to predict the subjective feelings of emotion (represented as arousal and valence) from the input of musical excerpts. The music excerpts are encoded as sound (psychoacoustic) features.

4.1 Simulations methodology

The experimental data for this simulation experiment, was obtained from a study conducted by Korhonen (2004a). The data is available online (Korhonen, 2004b), courtesy of the author. The original self report data includes the emotional appraisals of six selections of classical music, obtained from 35 participants (21 male and 14 female). Using a continuous measurement framework, emotion was represented by its valence and arousal dimensions (using the EmotionSpace Lab (Schubert, 1999b)). The emotional appraisal data was collected at 1Hz (second

by second).

4.1.1 Music pieces

Korhonen used six pieces of Western Art ("classical") music in his experiment (Korhonen, 2004a) (see Table 4.1). The pieces selected aimed to cover the widest range of emotion possible and to be musically diverse within the musical genre. The total duration of the pieces was limited to twenty minutes.

Piece ID	Title and Composer	Duration	Set
1	Concierto de Aranjuez - II. Adagio (J. Rodrigo)	165s	Training
2	Fanfare for the Common Man (A. Copland)	170s	Training
3	Moonlight Sonata - I. Adagio Sostenuto (L. Beethoven)	153s	Test
4	Peer Gynt Suite No 1 - I. Morning mood (E. Grieg)	164s	Training
5	Pizzicato Polka (J. Strauss)	151s	Test
6	Piano Concerto no.1 - I. Allegro maestoso (F. Liszt)	315s	Test

Table 4.1: Pieces used in Korhonen's experiment and their aliases for reference in this paper. The pieces were taken from Naxos's "Discover the Classsics" CD 8.550035-36

4.1.2 Psychoacoustic encoding (model input data)

Korhonen encoded the music pieces into the psychoacoustic space by extracting low and high level features, using Marsyas (Tzanetakis & Cook, 2000) and PsySound (Cabrera, 2000) software packages. Only Tempo was calculated manually, using Schubert's method as explained in Schubert (1999a)). The 11 psychoacoustic variables chosen (the 5 sound features representing Harmony variables included in Korhonen's study are not included here in order to exclude higher level features specific to the music culture and with controversial methods

for its quantification) are shown in Table 4.2 and described below (for convenience the input variables will be referred to with the aliases indicated in this table). Because some of these measures refer to the same psychoacoustic dimension, they were clustered into 6 major groups: Dynamics, Mean Pitch, Pitch Variation, Timbre, Tempo and Texture.

Sound feature	Group	Alias
Loudness Level	Dynamics	D_1
Short Term Maximum Loudness	Dynamics	D_2
Power Spectrum Centroid	Mean Pitch	P_1
Mean STFT Centroid	Mean Pitch	P_2
Mean STFT Flux	Pitch Variation	Pv_1
Standard Deviation STFT Centroid	Pitch Variation	Pv_2
Standard Deviation STFT Flux	Pitch Variation	Pv_3
Sharpness (Zwicker and Fastl)	Timbre	Ti_1
Timbral Width	Timbre	Ti_2
Mean STFT Rolloff	Timbre	Ti_3
Standard Deviation STFT Rolloff	Timbre	Ti_4
Beats per Minute	Tempo	T
Multiplicity	Texture	Tx

Table 4.2: Psychoacoustic variables considered for this study. The aliases indicated will be used in this article to refer to the variables they refer to in this table.

Dynamics: The Loudness Level (D_1) and the Short Term Maximum Loudness (D_2) represent the subjective impression of the intensity of a sound (measured in sones). Both algorithms estimate the same quantity (described in Cabrera, 1999) and output similar values.

Mean Pitch: The Mean Pitch was quantified using two power spectrum calculations (one from PsySound, and another from Marsyas). The Power Spectrum Centroid (P_1) represents the first moment of the power spectral density (PSD) (Cabrera, 1999). The Mean STFT Centroid (P_2) is a similar measure and corresponds to the balancing point of the spectrum (Tzanetakis & Cook, 2000) .

Pitch Variation: Contour was quantified using 3 measures. The Mean STFT Flux (C_1) corresponds to the Euclidian norm of the difference between the magnitude of the Short Time Fourier Transform (STFT) spectrum evaluated at two successive sound frames. The standard deviation of P_2 (C_2) and of C_1 (C_3), were also used to quantify the pitch variations[1] (refer to Tzanetakis & Cook, 2000 for furthers details).

Timbre: Timbre was represented using the 4 different measures: Sharpness (Ti_1), a measure of the weighted centroids of the specific loudness, approximates the subjective experience of a sound on a scale from dull to sharp - the unit of sharpness is the acum (one acum is defined as the sharpness of a band of noise centered on 1000 Hz, 1 critical-bandwidth wide, with a sound pressure level of 60 dB) - details on the algorithm used in Psysound can be found in Zwicker & Fastl, 1990); Timbral Width (Ti_2) is a measure proposed by Malloch (1997) which measures the flatness of the specific loudness function, quantified as the width of the peak of the specific loudness spectrum (see Cabrera, 1999 for further details and slight modifications to that algorithm); the mean and standard deviations of the Spectral Roll-off (the point where a frequency that is below some percentage of the power spectrum resides - refer to Tzanetakis & Cook, 2000 for the detail on these measures) are also two measures of spectral shape (Ti_3 and Ti_4) - although they do not directly represent timbre, Korhonen included these measures because they have been successfully used in music information retrieval.

[1] Although these algorithms are not specific measures of melodic contour, they have been successfully used as such in music information retrieval applications (Korhonen, 2004a). Nevertheless, in this article we refer to this variable as pitch variation because it characterizes better the nature of the encoding. Moreover, the relationships between pitch variations and emotion were the object of some studies (e.g. Scherer & Oshinsky, 1977), as described in Schubert (1999a).

Tempo: Tempo was estimated from the number of beats per minute. Because the beats were detected manually a linear interpolation between beats was used to transform the data into second by second values (details on the tempo estimation are described in Schubert, 1999a).

Texture: Multiplicity (Tx) is an estimate of the number of tones simultaneously noticed in a sound; this quantity was quantified using Parncutt's algorithm (as described in Parncutt (1989), page 92) included in Psysound.

4.1.3 Experimental data on subjective feelings (model output data)

Korhonen used the EmotionSpace Lab to quantify emotion on the dimensions valence and arousal (Schubert, 1999a). The emotional appraisal was collected at 1 Hz, when listeners were listening to each of the pieces. For this simulation experiment, the average arousal/valence values (for each second) over the 37 subjects were used to train the network.

4.1.4 Simulation procedure

The sound features constitute the input to the model. Each of these variables corresponds to a single input node of the network. The output layer consists of 2 nodes representing arousal and valence. Three pieces of music (1, 2 and 5), corresponding to 486s, were used during the training phase. In order to evaluate the response to novel stimuli, the remaining 3 pieces were used: 3, 4 and 6 (632s of music). Throughout this article, the collection of stimuli used to train the model will be referred to as "Training set", and "Test set" to the novel stimuli, unknown to the system during training, that test its generalisation capabilities and performance.

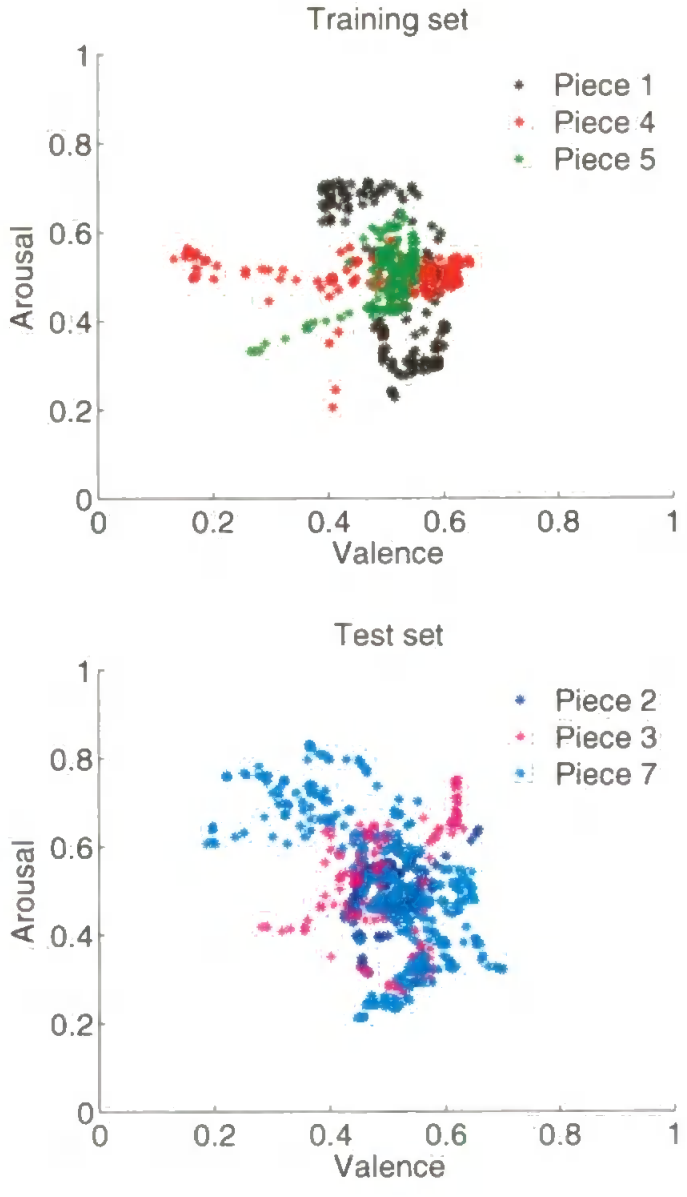


Figure 4.1: Areas covered by the pieces grouped by training (a) and test (b) sets. In both plots, the colours of each cloud of points correspond to all the arousal/valence pairs from each piece (see legend). Each point corresponds to the specific location on the 2DES of the arousal and valence values on a second by second basis.

The pieces were distributed among the sets in order to cover the widest range of values of the emotional space in both sets. The rationale behind the procedure is the fact that it is necessary to train the model with widest range of values possible in order to be able to predict the emotional responses to a diverse set of novel pieces. In ideal circumstances the areas covered by both sets should be similar. Figure 4.1 shows the areas of the 2DES covered by the pieces belonging to each data set and chosen following the procedure described. As it can be seen both sets contain extreme values in each variable and cover similar areas of the 2DES.

The task at each training iteration (t) is to predict the next ($t + 1$) values of arousal and valence. The target values (aka "teaching input") are the average arousal/valence pairs across all participants in Korhonen's experiments. In order to adapt the range of values of each variable to be used with the network, all variables were normalised to a range between 0 and 1. Figure 4.2 shows the general neural network architecture. In each simulation experiment the exact number of inputs, hidden and memory units will be specified, as these depend on the input layer configuration for selection of acoustic features.

The learning process was implemented using a standard backpropagation technique (Rumelhart et al., 1986). During training the same learning rate and momentum were used for each of the 3 connection matrices. The network weights were initialised with different random values. The range of values for each connection in the network (except for the connections from the hidden to the memory layer which are set constant to 1.0) was defined randomly between -0.05 and 0.05.

If the model is also able to respond with low error to novel stimuli, then the training algorithm was able to extract some general rules from the training set that relate musical features to emotional ratings. The maximum number of training iterations and the values of the learning parameters were estimated in order to

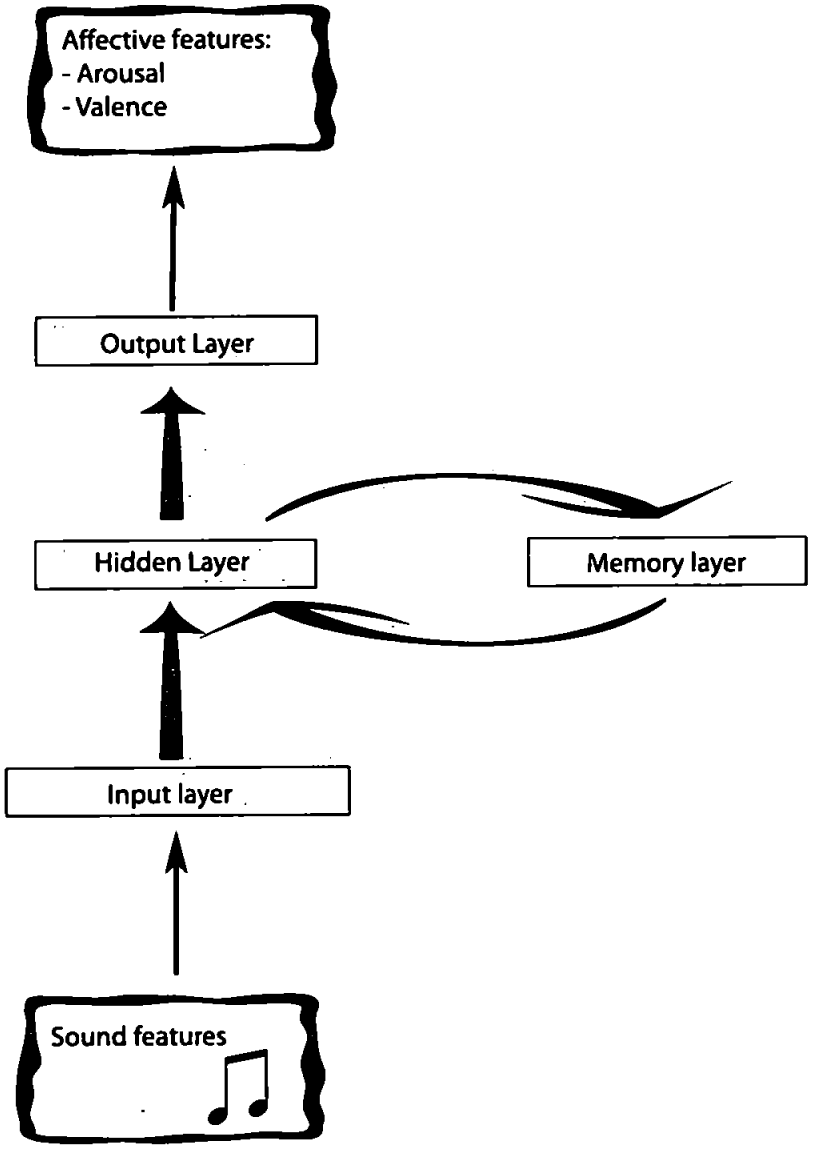


Figure 4.2: Schematic of an Elman neural network as used in simulations. It consists on a variation on the multilayer perceptron, with an extra memory layer, which provides the network with feedback connections. The inputs to the network are the sound (psychoacoustic) features, and the outputs the affective dimensions under investigation (arousal and valence).

avoid the overfitting of the training set data[2]. After preliminary tests and analysis, the duration of the training was set at 20000 iterations, using a learning rate

[2]In order for the model to generalise, it must not be built around the minimisation of the error in the training data. The ideal point is a compromise between the output errors for both training and test data.

of 0.075 and a momentum of 0.0. Testing the model with different numbers of hidden nodes also permitted the optimisation of the size of the hidden layer, which defines the dimensionality of the internal space of representations. The best performance was obtained with a hidden layer of size 5.

The root mean square (rms) error is used here to quantify the differences between values predicted by the model and the values actually observed experimentally. Although this is a common measure to assess the performance of connectionist models, it does not assure a successful modelling process. The purpose of using this measure is only to compare the model performance with alternative sets of inputs to the network. To assess the model ability to classify the stimuli in terms of their affective value (and so the effectiveness of the model), the model categorisation process will also be analysed in detail.

4.2 Reduction of the psychoacoustic (input) dimensions

The choice of the sound features must consider musical, psychological and modelling aspects. The sound features chosen by Korhonen include a significant set of perceptually relevant dimensions, although there are some redundancies to address. A recurrent problem in dealing with this type of data are the correlations among the encoded dimensions, especially redundant information and collinearity (as discussed by Schubert (1999a)). To avoid this, only one variable from each psychoacoustic dimension will be used.

4.2.1 Testing individual sound features

Tempo, Texture, Dynamics, Mean Pitch, Pitch Variation, and Timbre are all considered to be included in the model as separate dimensions. In the case of Tempo and Texture, because they ere estimated using a single method (algorithm), they are included directly because theres no choice among

alternative measures to be made[3]. In order to select one sound feature from the remaining musical dimensions (Dynamics, Mean Pitch, Pitch Variation, and Timbre), each set of inputs considered included all unique features for each musical dimension as a basic set (tempo and texture as explained before), plus one other test variable(s). For instance, in the case of Dynamics[4] T, Tx, D_1 and D_2 were tested, but also T, Tx and D_1, and T, Tx and D_2. The same procedure was followed for Mean Pitch, Pitch Variation and Timbre.

For each test case I did 3 simulation replications, i.e. trained 3 different neural networks (with different random configuration of initial weights) and averaged their errors. Table 4.3 shows the rms errors for each test condition. These results are used to select one sound feature per psychoacoustic dimension (loudness, timbre, mean pitch and pitch variation).

For comparison purposes, the mean RMS error of both outputs of the network, before learning (i.e. with random weights), to all pieces was calculated: $rms_{random} = 0.107$. It corresponds to the error value when all the pieces are fed to a model with random weights. The error of each simulation test case was compared with this value. For all psychoacoustic (sound) dimensions, at least one variable improves the model performance ($rms_{testcase} < rms_{random}$).

For the loudness measures, it was found that the inclusion of both variables ($rms = 0.066$) whilst the sole inclusion of the loudness level measure ($rms(D_1) = 0.067$) produced similar results. Therefore the loudness level (D_1) was the key variable selected from this group. Regarding timbre, the best performance was achieved using only Zwicker and Fastl's (1990) sharpness measure ($rms(Ti_1) = 0.077$), and so this key variable was also selected. The key variable selected to represent the mean pitch level is the power spectrum centroid ($rms(P_1) = 0.082$),

[3]Tempo (T) and Texture (Tx) were chosen as the variables for the initial features for several reasons. First, is that they are the only variable for the sound features that they represent. A second important factor is that tempo and texture are expected to contain important information about changes in the affective experience (Schubert, 1999a).

[4]In Table 4.3, the index "all" is used when all variables from a specific sound feature are included; for instance, D_{all} indicates the inclusion of D_1 and D_2.

Test group	Input features	rms Train A	rms Train V	rms Test A	rms Test V	rms av.
Loudness	T-Tx-$\mathbf{D_{all}}$	0.056	0.061	0.068	0.080	**0.066**
	T-Tx-$\mathbf{D_1}$	0.058	0.058	0.072	0.081	**0.067**
	T-Tx-D_2	0.066	0.056	0.088	0.077	0.072
Timbre	T-Tx-Ti_{all}	0.074	0.067	0.088	0.087	0.079
	T-Tx-$\mathbf{Ti_1}$	0.069	0.063	0.082	0.093	**0.077**
	T-Tx-Ti_2	0.105	0.073	0.098	0.080	0.089
	T-Tx-Ti_3	0.108	0.074	0.135	0.121	0.110
	T-Tx-Ti_4	0.110	0.076	0.130	0.087	0.101
Mean pitch	T-Tx-P_{all}	0.080	0.072	0.106	0.095	0.088
	T-Tx-$\mathbf{P_1}$	0.072	0.066	0.107	0.083	**0.082**
	T-Tx-P_2	0.136	0.083	0.233	0.106	0.140
Pitch variation	T-Tx-$\mathbf{Pv_{all}}$	0.100	0.062	0.119	0.083	**0.091**
	T-Tx-$\mathbf{Pv_1}$	0.101	0.064	0.130	0.086	**0.095**
	T-Tx-Pv_2	0.108	0.070	0.134	0.083	0.099
	T-Tx-Pv_3	0.102	0.067	0.133	0.090	0.098
All	T-Tx-D_1-Ti_1-P_1-Pv_1	0.048	0.049	0.068	0.076	**0.060**

Table 4.3: rms error for each input data set using a model with 5 hidden units. The values shown were averaged across 3 replications of each simulation test case. For comparison purposes, the mean rms error of both outputs for a network with random weights was was established as reference value ($rms_{random} = 0.107$).

because it performs better than the remaining variables. Finally, the pitch variation features show very similar error values for all test cases. The mean STFT flux was the key variable chosen because it yields a lower error ($rms(Pv_1) = 0.095$) than the standard deviation of the STFT centroid ($rms(Pv_2) = 0.099$) and the standard deviation of the STFT flux ($rms(Pv_3) = 0.098$).

4.2.2 Selected sound features: model inputs

In all subsequent simulations I will use an ENN network that includes all the variables chosen above (T, Tx, D_1, Ti_1, P_1 and Pv_1), in order to assess the performance with all variables together. The results of this combination of inputs are shown in the final row of Table 4.3. An inspection of the rms error shows that combining all these features improved the model's performance substantially, suggesting that the interaction among different features conveys

relevant information. In the following simulation experiment, these 6 key sound features will be used as the inputs to the model. This model will be referred to as "Model 0" throughout this thesis. Its architecture is shown in Fig. 4.3.

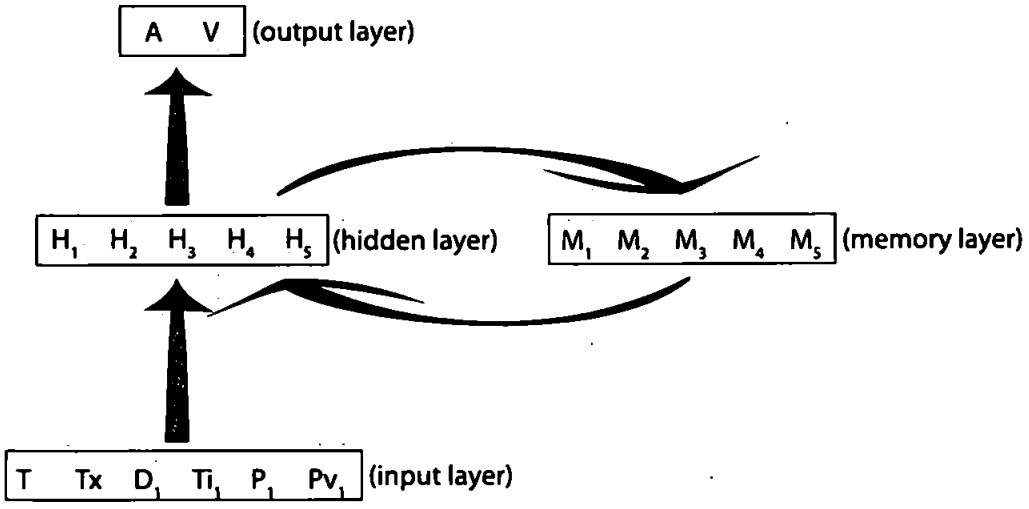


Figure 4.3: Neural network architecture and units identification for Model 0 (model used in simulation experiments): Input units - sound features (T, Tx, L, P, S and Pv); Hidden units - H_1 to H_5; Memory (context) units - M_1 to M_5; Output units - arousal (A) and valence (V).

4.3 Analysis of model performance

After choosing the set of input features to be used with the model, 37 neural networks (the same as the number of participants in Korhonen experiments) were trained using the network configuration as in Fig. 4.3). The average error (for both outputs) of the 37 networks was 0.050 for the Training set, and 0.076 for the Test set. These values were produced after 20000 iterations of the training algorithm.

In order to compare the model output with the experimental data for each piece, the rms error and the linear correlation coefficient (r) were used to describe the deviation and similarity between the model outputs and the experimental data.

The following analysis will report on the network that showed the lowest average error for both data sets (network 24). The rms error and r of each output for all the pieces are shown in Table 4.4.

Piece ID	rms error		r		Set
	A	V	A	V	
1	0.052	0.044	0.964*	0.760*	Train
2	0.040	0.054	0.778*	0.939*	Train
3	0.061	0.045	0.278**	0.206***	Test
4	0.085	0.081	0.797*	0.040	Test
5	0.044	0.046	0.768*	0.583*	Train
6	0.052	0.082	0.958*	0.650*	Test
av.	0.056	0.059	0.757	0.530	

Table 4.4: Comparison between the model outputs and experimental data: root mean square (rms) error and linear correlation coefficient (r) ($^{*}p < 0.0001$, $^{**}p < 0.001$, $^{***}p < 0.02$)

Figures 4.4 and 4.5 show the arousal and valence outputs of the model for Training and Test sets, versus the data obtained experimentally (target values).

The model was able to track the general fluctuations in arousal and valence for both data sets, whilst the performance varied from piece to piece. The model performance for arousal was better for pieces 1, 2, 5 and 6 ($rms_1 = 0.052$, $rms_2 = 0.040$, $rms_5 = 0.044$ and $rms_6 = 0.052$), as shown by the low rms errors (lower than the mean arousal for all pieces: $rms_{all} = 0.059$) and high r.

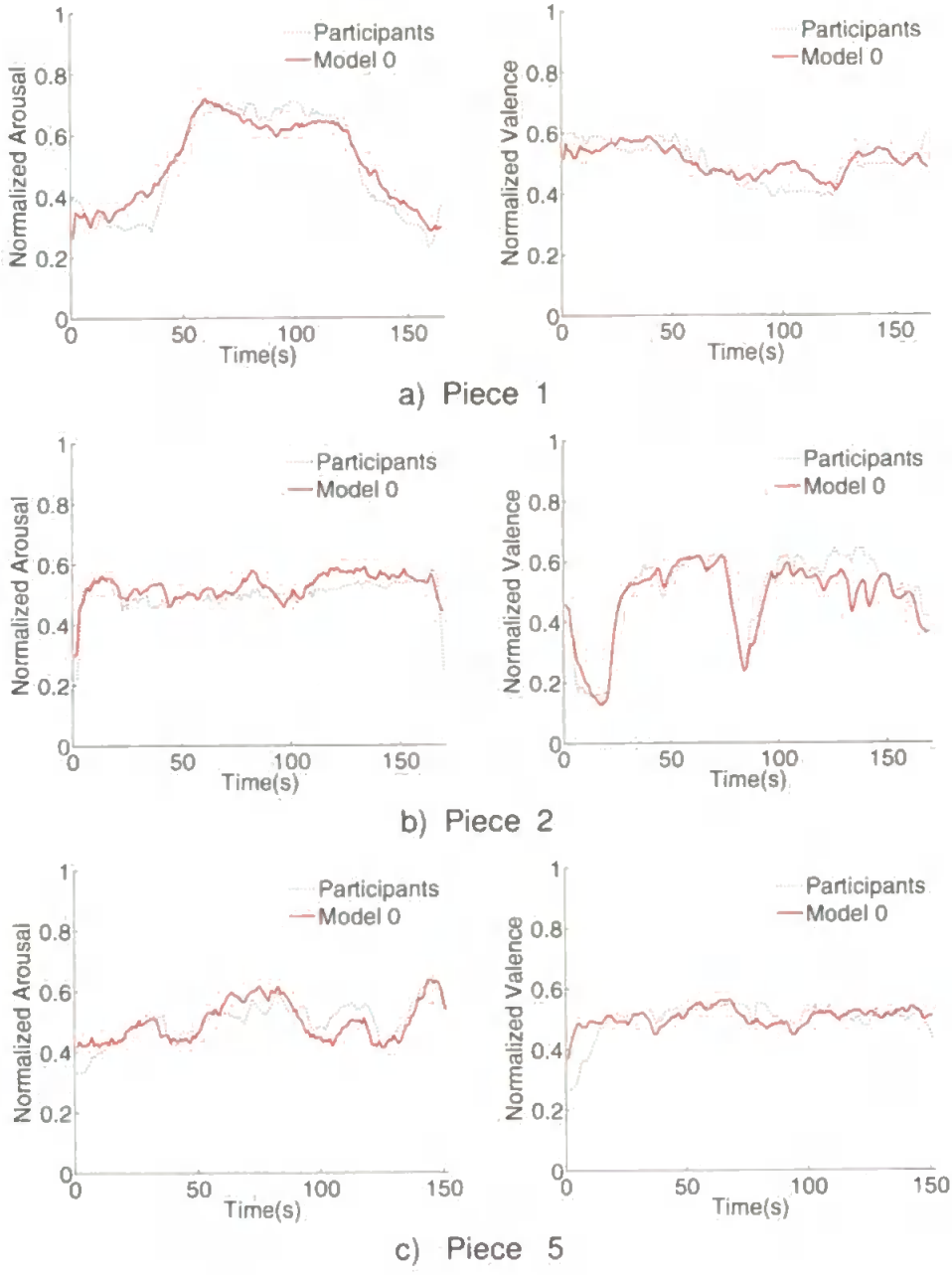


Figure 4.4: Training pieces - Arousal and Valence model outputs compared to experimental data for the training data set: a) Piece 1 (Rodrigo, *Concierto de Aranjuez*), b) Piece 2 (Copland, *Fanfare for the Common Man*) and c) Piece 5 (Strauss, *Pizzicato Polka*).

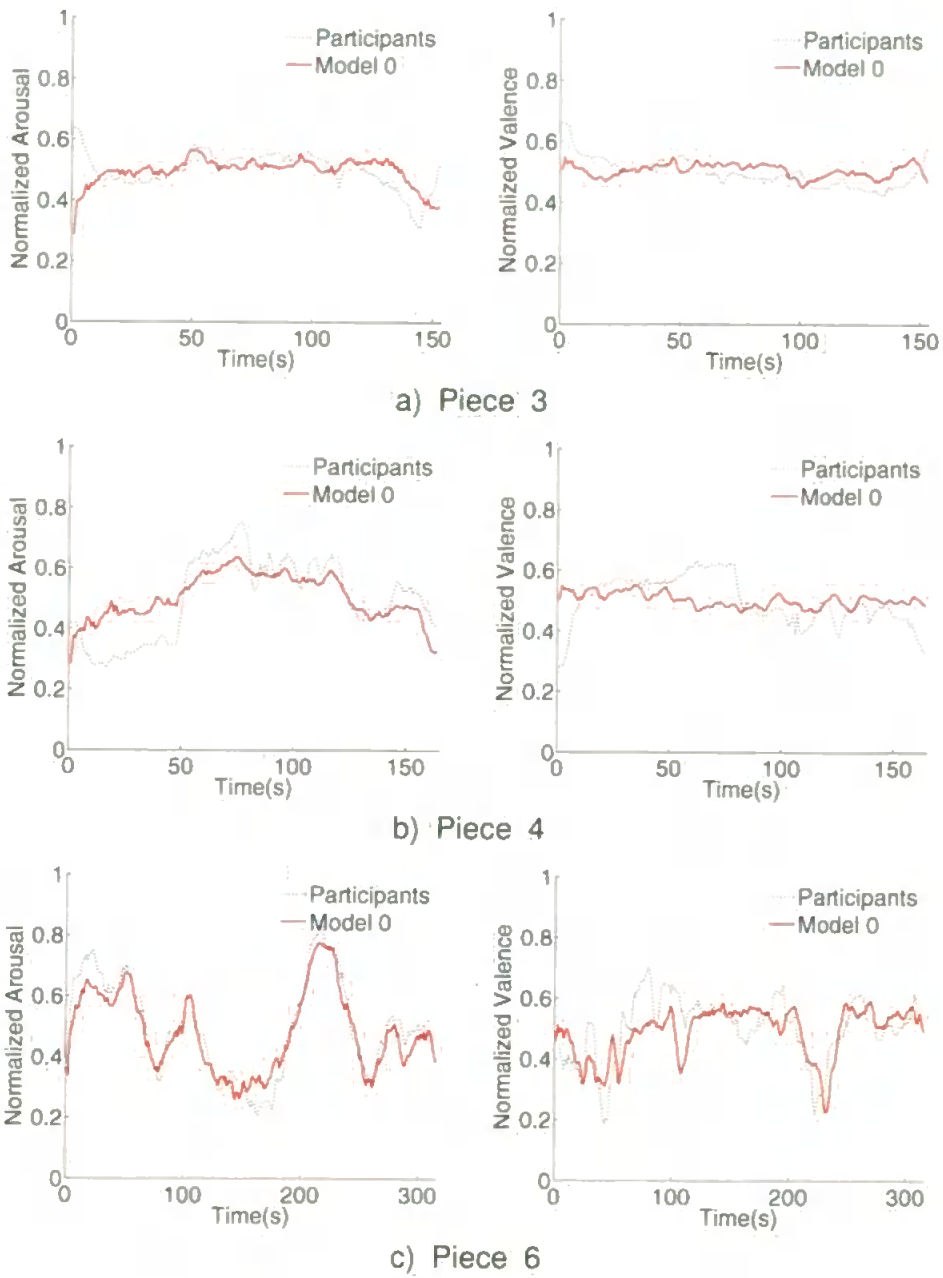


Figure 4.5: Test pieces - Arousal and Valence model outputs compared to experimental data for the test data set: a) Piece 3 (Beethoven, *Moonlight Sonata*), b) Piece 4 (Grieg, *Peer Gynt Suite No 1*) and c) Piece 6 (Liszt, *Piano Concerto no.1*).

Pieces 3 and 4 had a higher rms error than the mean of all the remaining pieces. While piece 3 has low correlations coefficients with both arousal ($r = 0.278$) and valence ($r = 0.206$) relatively to the rest of the pieces, piece 4 has no significative correlation coefficient for valence ($r = 0.040$). This weaker performance is visible in Figure 4.5 (a) and b)). The initial 80s (approximately) of the model predictions show different patterns: the model predictions have a decreasing tendency while the experimental data shows an increase). The error is particularly evident after the initial section (that last for 50s, a "dialogue" between flutes and strings) to which follows a strong increase in valence (only strings playing in bigger number louder) until around 80s of the piece (transition to a new section in piece).

The overall successful predictions of the affective dimensions for both known and novel music support the idea that music features contain relevant relationships with emotional appraisals. A visual inspection of the model outputs, confirmed by the rms and r measures, also indicate that the model output resembles the experimental data. The spatiotemporal relationships learned from the Training set were successfully applied to a new set of stimuli.

These relationships now encoded in the network weights, and the flux of information in the internal (hidden) layer of the neural network represents the dynamics of the internal categorisation (or recombination) of the input stimuli, that enables output predictions. One of the advantages of working with an artificial neural network is the ability to explore the internal mechanisms, which generate the behaviour and indirectly show how the model processes the information.

To study the relationships between the sound features and the model's predictions it is necessary to identify how the hidden units process the inputs into the outputs. With that information it is possible to estimate the input-output transformations of the model. One possibility is to inspect the weights matrices in the model and identify the highest weights. Although simple, this methodology

focuses on the long-term memory of the model, which totally discards the dynamics of the model: the temporal structure of the data flow in the system, a kind of "task specific" memory.

In order to analyse both the long-term and short-term temporal dynamics of the model, two different methods are used: lesioning and correlation analysis. The first method consists of removing the connections between successive pairs of processing units in the model. By doing this, it is possible to identify which hidden units relate to the inputs (or any other connected unit), and compare their relative contribution to the model predictions. The second method provides a measure of the correlation between the activity of the different layers in the network, by quantifying the overall (spatial and temporal) relationships between groups of sound features (inputs), hidden units and outputs (arousal and valence).

In the following section, analysis of the model is conducted in three steps. The first step focuses on the analysis of the task specific representations of the model (temporal structure). This will reveal the way in which the network represents music inputs in order to predict changes in their affective value. Then, in the second step, the relative contribution of each input to the internal representations (hidden units activity) is quantified, based on the weights matrices analysis. To complete the analysis procedure, step 3, the temporal correlations between inputs, hidden units and outputs are analysed.

4.3.1 Model internal dynamics: discriminant functions

In order to analyse the internal dynamics of the model the Linear Discriminant Analysis (LDA) was used (McLachlan, 1992). LDA is a classical method of classification using categorical target variables (features that somehow relate to or describe objects). Unlike Principle Component Analysis (PCA), in LDA the

groups are known or predetermined[5]. The main purpose of this algorithm is to find the linear combination of features which best separate classes or object properties. This method maximises the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. The classification model chosen consists of the 4 quadrants of the two-dimensional emotional space (2DES) (Q_1, Q_2, Q_3 and Q_4). Because the hidden units allow the model to develop representations that are selected by the constraints of the task, the hypothesis is that the of the A/V space structure represents the underlying internal space of representation in the model. This method also allows us to identify the hidden units related to each dimension of the categorical space (an important aspect because it may facilitate the investigation of the input-output mapping of the model).

The analysis has shown that 2 discriminant functions can explain 99.7% of the variance in the data[6]. The canonical correlations of the original data set are 0.821 for the 1^{st} discriminant function (F_1) and 0.506 for the 2^{nd} function (F_2). In Fig. 4.6, we show the 2 discriminant functions. Each point corresponds to an internal state of the model. The colours and numbers identify the internal states of the model belonging to each of the categories hypothesised (the affective space quadrants).

The model shows an internal discrimination of the input stimuli very similar to the affective space quadrants division. This indicates that the input stimuli were successfully categorised according to their affective value. As the discriminative power of the model is embedded in the hidden units activations (the ones that connect to the output), it is necessary an assessment of the influence of each hidden unit on the pair of canonical variables needs. This was done by analysing

[5]Both methods are very similar because they look for linear combinations of variables which best explain the data; the essential difference consists of the rules for classification (or clustering), which is based on distance measures in PCA while LDA explicitly attempts to model the difference between the classes.

[6]This does not mean that we can reduce the number of units in the model, but instead that some of these units might vary along similar dimensions. As it will be shown, all the hidden units have relevant contributions to at least one of the discriminant functions

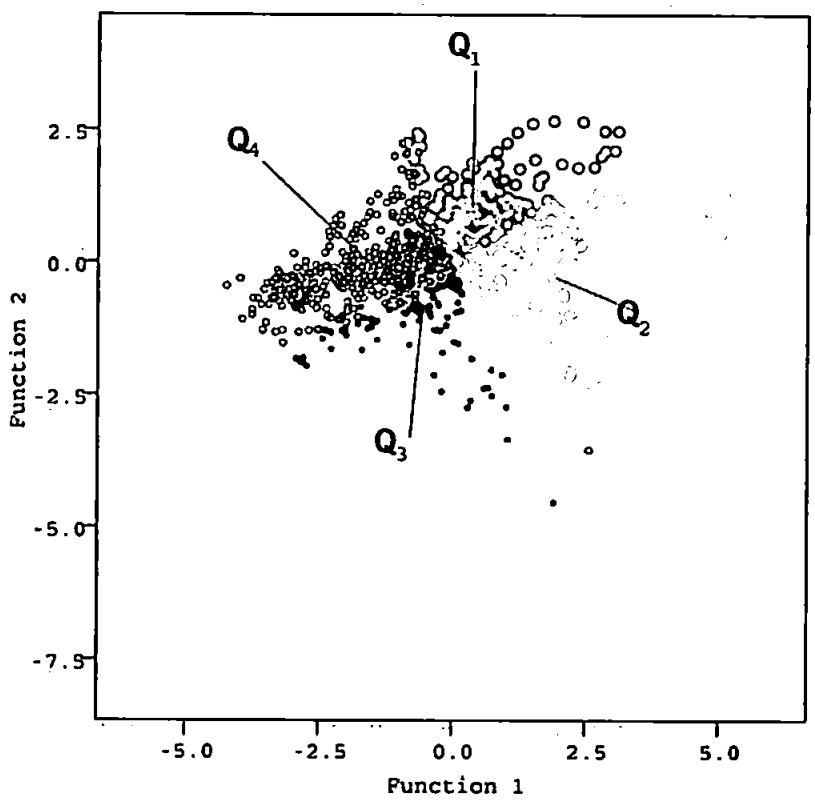


Figure 4.6: Linear discriminant analysis of hidden layer activations (canonical discriminant functions plot). The labels Q_1, Q_2, Q_3 and Q_4 indicate the quadrant in the 2DES that each coloured cluster represents. The plot shows that the spatiotemporal structures detected by the model are organised as a 2-dimensional space, and differentiated in terms of arousal and valence.

the factor structure coefficients shown in Table 4.5. These values correspond to the correlations between the variables in the model and each of the discriminant functions (similar to the factor loadings of the variables on each discriminant function in PCA).

Hidden unit	F_1	F_2
H_1	-0.489	-0.246
H_2	0.371	0.896
H_3	-0.788	0.291
H_4	-0.520	0.569
H_5	0.633	-0.604

Table 4.5: Factor Structure Matrix: correlations between discriminant variables and each hidden unit.

The 1^{st} discriminant function (F_1) receives the highest contributions from H_1, H_3, H_4 and H_5. F_2 receives the strongest contributions from H_2, H_4 and H_5. The correlations between the discriminant functions and the hidden units can symbolically be represented as: $F_1 = -H_3 + (H_5 - H_4)$ and $F_2 = H_2 - (H_5 - H_4)$. In order to study the relationships between sound features and model predictions, the following step consists of identifying how the hidden units activity (or the internal model states) relate to the input and output spatiotemporal dynamics. This is achieved through lesion analysis.

4.3.2 Lesioning tests: long-term memory analysis

Before applying the lesioning tests, the first step is to analyse the weight matrices between the different layers. The weights are represented in Figure 4.7 as rectangles of variable size and colour. The size is proportional to the weight value (bigger rectangle, bigger weight), and the colour represents the signal of the weight - red for negative, green for positive. All three learned weight matrices in the model are represented (note that weights from hidden to memory layer are kept constant).

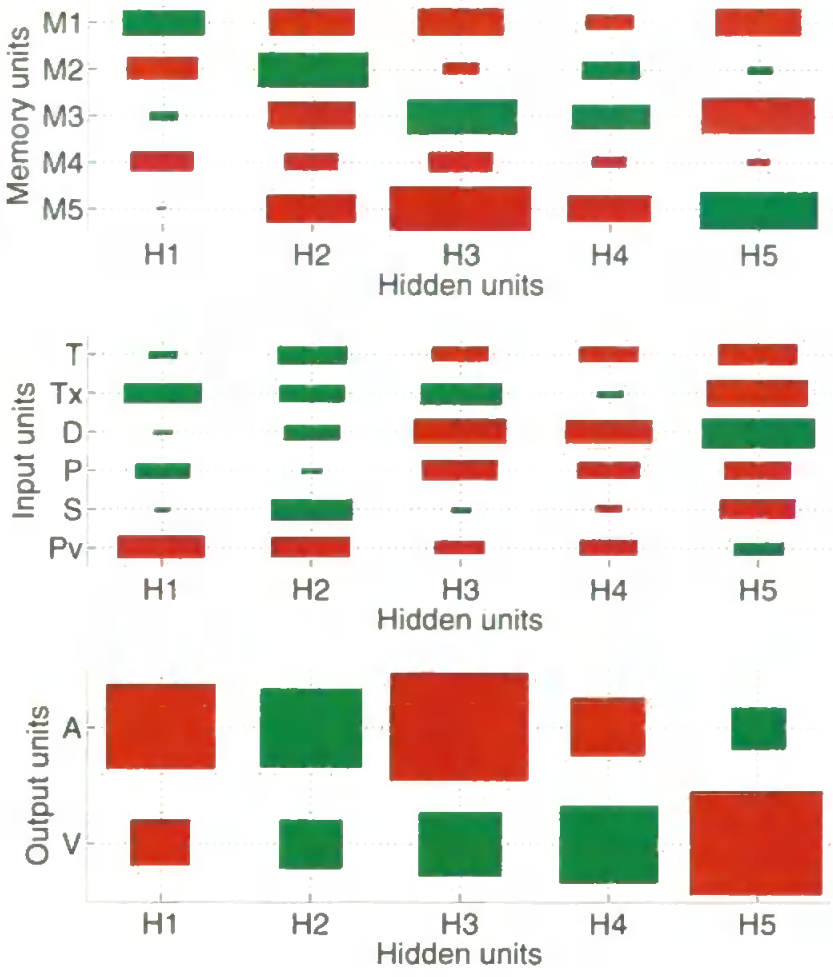


Figure 4.7: Neural network weight connection matrices: memory to hidden layers (top), input to hidden layers (middle) and hidden to output layers (bottom). The weights are represented as rectangles of variable size and colour: the size is proportional to the weight value and the colour represents the signal of the weight (red for negative and green for positive).

The network contains a diverse range of different weights magnitudes with both inhibitory and excitatory connections. This suggests that the network can distinguish inputs with small differences. The same is also true for their temporal structure, which indicates the important role of the temporal structure of the features extracted from the inputs (the compressed representation of past states)

to predict the outputs (the standard deviation of the weights from the input to the hidden layer is 1.30, and from the memory layer to the hidden layer is 1.90). There is also a tendency for the memory units to reinforce their respective hidden units (with the exception of H_4), and to inhibit other hidden dimensions. Some more information can be obtained from the lesioning tests.

The lesioning tests consist of systematic "damaging" of selected connection weights in the model. The variations of the output error (e.g increase) gives an idea of how important this connection is to the model output dynamics. The lesioning procedure applied here involves lesioning each of the connections of all 3 learned weight matrices. In order to simplify the visualisation of the lesioning analysis[7], Figure 4.8 shows the effect of removing each of the weights from the network on the output predictions. If the removed connection had little effect on the output performance ($rms < 0.090$), it is represented in black. This means low rms error, similar to the post training error (unlesioned network). Higher values of error are represented in gray ($0.090 < rms < 0.300$) and white ($rms >= 0.300$).

The lesioning tests shows two distinct groups of hidden units with stronger effects on the output: H_1 ($rms = 0.231$), H_2 ($rms = 0.225$) and H_3 ($rms = 0.181$) are related to arousal, while H_2 ($rms = 0.111$), H_4 ($rms = 0.125$) and H_5 ($rms = 0.159$) are related to valence. Since the hidden units are the only units directly connected to the output, these relationships are mirrored in the weight connections between these two layers.

All the sound features have relevant information for the model: for all inputs, at least two damaged connections to the hidden layer decreased substantially the model performance (for arousal and valence). H_4 is the only hidden unit which blocks the input information. For both outputs, the removal of each connection from the input layer to this unit had a small effect on the output when removed (all

[7]The complete lesioning procedure involved 60 tests. As it would be difficult to analyse all this charts separately, the following diagram aims to integrate all the results into a single representation scheme.

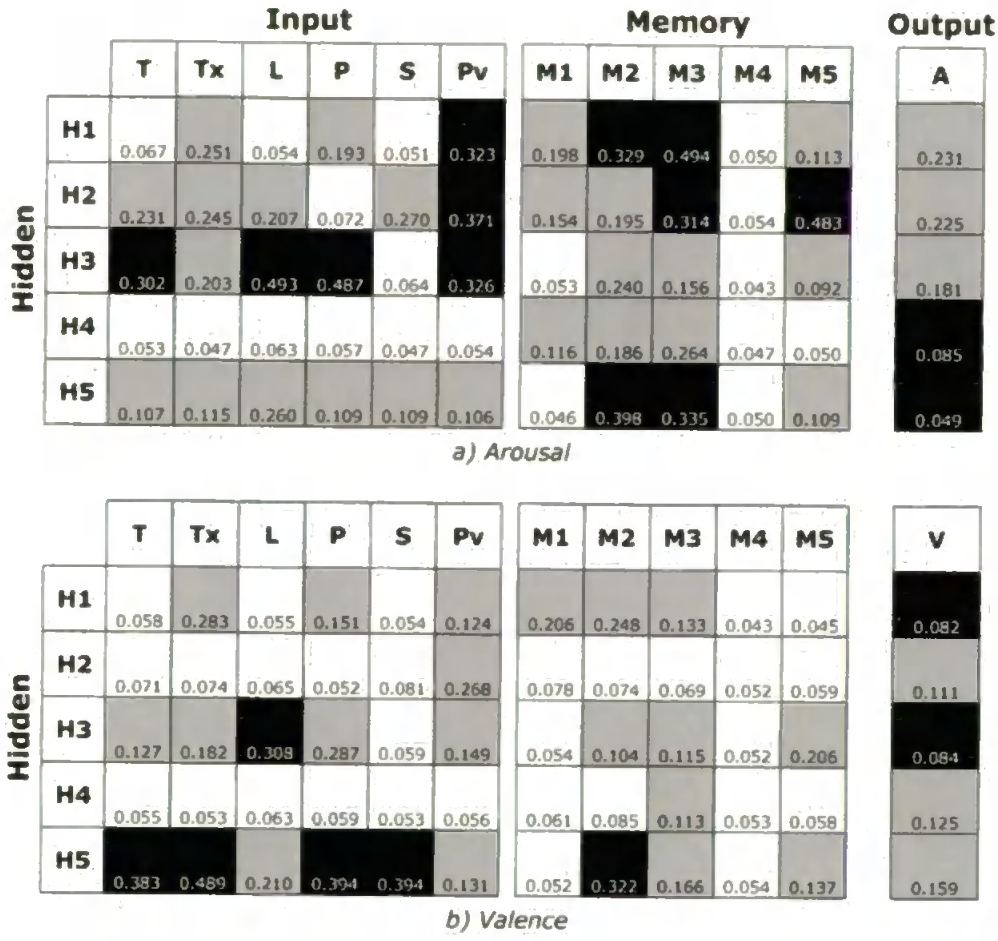


Figure 4.8: Model weight matrices analysis: each learned weight in the model was removed (value set to 0.0, one at a time), and the model performance was then measured using the rms error. Each cell in the above matrices corresponds to the removal of one connection linking two processing units, and the values indicated in each cell correspond to the rms error. For easier reading, the rms errors are represented using a colour code: black for those weights that had small or no effect on the model performance ($rms < 0.09$); for higher errors grey ($0.09 < rms < 0.30$) and white were used ($rms >= 0.30$).

rms were lower than 0.063). Unlike the lesioning tests on the hidden to output units, the input to the hidden layer connections are not the only units affecting the hidden layer activity. The compact representation of the past states of the network is sent to the hidden layer from the memory layer. These are the connections that "decode" the temporal structure of the inputs.

As expected, the temporal structure of the sound features was fundamental

to the prediction of arousal and valence. The first main observation is that the arousal output had a stronger attachment with the memory layer: 17 out of 25 of the weights in this matrix had high errors when removed. Valence had only 10 relevant connections. M_4 is the only memory unit with very small interference in the hidden layer activity (the effect of removing each connection had very little effect on the modem performance; the rms was always lower than 0.054) and the output predictions. Because H_4 is isolated from the inputs and its past state (M_4) is discarded by the model, its activity is only affected by the past states of hidden units 1, 2 and 3 (which as seen are related to the arousal output). The recombination of past states in H_4 has nonetheless a relevant connection to the valence output. Because the activity in H_4 is related to the arousal context (H_1, H_2 and H_3), the valence predictions have commonalities with the arousal predictions.

4.3.3 Input/output transformation: model production rules

The above observations quantify the interaction between the different processing units. They define the long-term memory of the model that permits the spatiotemporal differentiation of the inputs stimuli. However, to study the relationships between inputs and model predictions it is also required to analyse the data flow within the model. Due to the recurrent links, the input information can be recombined and reflected in the activity of different hidden units. In order to gain insight about the inputs propagation through the model, it is necessary to use a measure of shared activity in the input and hidden layers. Sometimes, the weights between two units are of small magnitude, but the respective correlation between their activities is high. The opposite may also occur.

While the weights pertain to the unique contribution of each variable, the correlations between the units' activity represent the overall contribution. For example, if two variables (such as loudness and texture), contain redundant information, then the model may rely less on one of them, because it only

"needs" to include one of the items to capture the essence of what they measure. Once a large weight is assigned to one of the variables, the contribution of the second item may be redundant and, consequently, it will receive a smaller (or even negligibly) small weight (for instance, Texture is many times related to the loudness level since more notes sounding or instruments playing frequently relate with increased loudness). Nevertheless, by looking at the correlations between the inputs and the hidden units' activity, those may be substantial for both. To reiterate, the weights pertain to the unique contributions of the respective variables with a particular weighted sum (or processing unit activity).

In order to account for the temporal dynamics of the model, the correlations between inputs, hidden and output units were computed using a Canonical Correlation Analysis (CCA) (Hotelling, 1936). A canonical correlation is the correlation of two canonical variables: one representing a set of independent variables, the other a set of dependent variables. The CCA optimises the linear correlation between the two canonical variables to be maximised in the context of many to many relationships. There may be more than one linear correlation relating the two sets of variables, each representing a different dimension of the relationship, which explain the relation between them. For each dimension it is also possible to assess how strongly it relates each variable in its own set (canonical factor loadings). These are the correlations between the canonical variables and each variable in the original data sets. While the lesioning tests facilitated the selection of groups of hidden units with strong relationships to the output, this analysis aims at investigating the model's internal dynamics and its correlations with the inputs and outputs.

The CCA is used to assess the relationships between the sequences of input, hidden and output layers activity. This method permits the analysis of the contribution of each network layer node or (sets of nodes) to the activity of a different layer. Relevant for the analysis are the relationships between input

and hidden layers (how the inputs relate with the internal representations of the model), and these with the outputs (which sets of hidden units are more related to the output). In Table 4.6 the details of a CCA for the activity of the neural network layers are shown.

Loadings (Input/Hidden)			
Variable	var. 1	var. 2	var. 3
H_1	-0.398	**-0.633**	-0.028
H_2	**0.479**	**0.657**	-0.437
H_3	0.144	**-0.891**	-0.238
H_4	0.159	**-0.647**	-0.632
H_5	**-0.637**	**0.645**	0.018
T	0.264	0.478	0.151
Tx	**0.608**	0.280	0.217
D	0.450	**0.674**	0.139
P	**0.819**	0.297	0.432
Ti	**0.748**	0.420	0.262
C	0.187	0.270	0.825
Canon Cor.	**0.725**	**0.546**	0.448
Pct.	**61.1%**	**23.4%**	13.8%

Loadings (Hidden/Output)		
Variable	var. 1	var. 2
H_1	**-0.504**	0.482
H_2	**0.978**	-0.055
H_3	-0.291	**0.862**
H_4	0.014	**0.797**
H_5	-0.074	**-0.973**
A	**0.765**	**-0.644**
V	0.260	**0.966**
Canon Cor.	**0.987**	**0.984**
Pct.	**66.0%**	**44.0%**

Table 4.6: Canonical Correlation Analysis (CCA): the canonical correlations (the canonical correlations are interpreted in the same way as the Pearson's linear correlation coefficient) quantify the strength of relationships between the extracted canonical variates, and so the significance of the relationship. To assess the relationship between the original variables (input, hidden and output units activity) and the canonical variables, the canonical loadings (the correlations between the canonical variates and the variables in each set) are also included.

The bigger the loading, the strongest relationships between the original variables (input, hidden, and output units' activity) and the canonical variates. The following paragraphs summarise these relationships, which explain how the network inputs are propagated through the hidden layer to the network's outputs.

Input to hidden: Three canonical variables explain 98.3% of the variance in the data (see left side of Table 4.6). The first pair of variables loads on P, Tx, Ti (inputs set), H_2 and H_5 (hidden layer). The second, loads only on input D, but it loads on all nodes of the hidden layer. The third canonical variable loads on C,

H_2 and H_4. These 3 dimensions encode the general levels of shared activation in the input and hidden layers.

Hidden to output: Two canonical variables explain all the variance in the data (see right side of Table 4.6). The first root is correlated strongly with arousal, and the activity in hidden units H_1 and H_2. The second pair of canonical variables correlates with both valence (positive) and arousal (negative), and with the activity in units H_3 to H_5.

By taking these 2 groups of relationships together it is possible to establish qualitative patterns of correlations illustrative of the general model dynamics. The lesioning tests facilitated the selection of groups of hidden units with strong relationships to the output, while the CCA has shown how the model's internal dynamics correlates with the inputs. By using both analyses the output units' activity can, symbolically, be represented as[8]:

$$A(t) = g_1(-H_1(t), H_2(t), -H_3(t)) + m_1(M_1(t), M_2(t), M_3(t), M_5(t)) \quad (4.1)$$

$$A(t) = g_1(T(t), Tx(t), L(t), P(t), S(t)) + m_1(H_1(t-1), H_2(t-1), H_3(t-1), H_5(t-1)) \quad (4.2)$$

At a given time (t), arousal ($A(t)$) is positively associated with the current T, Tx, L, P and S inputs, plus the memory of previous states (except for the H_4 dimension). Applying the same principle to valence leads to:

$$V(t) = g_2(-H_2(t), H_4(t), -H_5(t)) + m_2(M_3(t), M_5(t)) \quad (4.3)$$

$$V(t) = g_2(-H_2(t), H_4(t), -H_5(t)) + m_2(H_3(t-1), H_5(t-1)) \quad (4.4)$$

[8]The signal of the canonical loadings indicates if a hidden unit reinforces or inhibits the outputs. The hidden units considered correspond to the fundamental units to the output predictions found in the lesioning analysis, while the input units correspond to the strongest correlations with the hidden layer found with CCA.

Considering that H_4 blocks all the inputs (this unit is only affected by the memory layer units), and that H_2 was almost linearly related to arousal[9], we can further simplify:

$$V(t) = g_2(-H_5(t)) + m_2(H_1(t-1), H_2(t-1), H_3(t-1), H_5(t-1)) \quad (4.5)$$

$$V(t) = g_2(T(t), -L(t), P(t), S(t)) + m_2(H_1(t-1), H_2(t-1), H_3(t-1), H_5(t-1)) \quad (4.6)$$

4.3.4 Summary

Figure 4.9 provides a qualitative representation of the relationships between sound features and affective dimensions. This representation is merely representative of the main observable fluxes of information in the model. They do not represent the complete interaction between inputs and outputs, but they give a general overview of contributions.

The general strategies for input-output (sound features - affective dimensions) mapping found are:

Tempo (bpm): fast tempi are related to high arousal (quadrants 1 and 2), and positive valence (quadrants 1 and 4). Slow tempi exhibit the opposite pattern;

Multiplicity (texture): thicker textures have positive relationships with arousal (quadrants 1 and 2);

Loudness (dynamics): higher loudness relates to increased arousal and decreased valence;

[9]The canonical loadings associated with the first canonical function and H_2 are 0.978 for arousal and -0.055 for valence; because of that, in order to consider only the effect on valence, the parameter was omitted in the symbolic representation

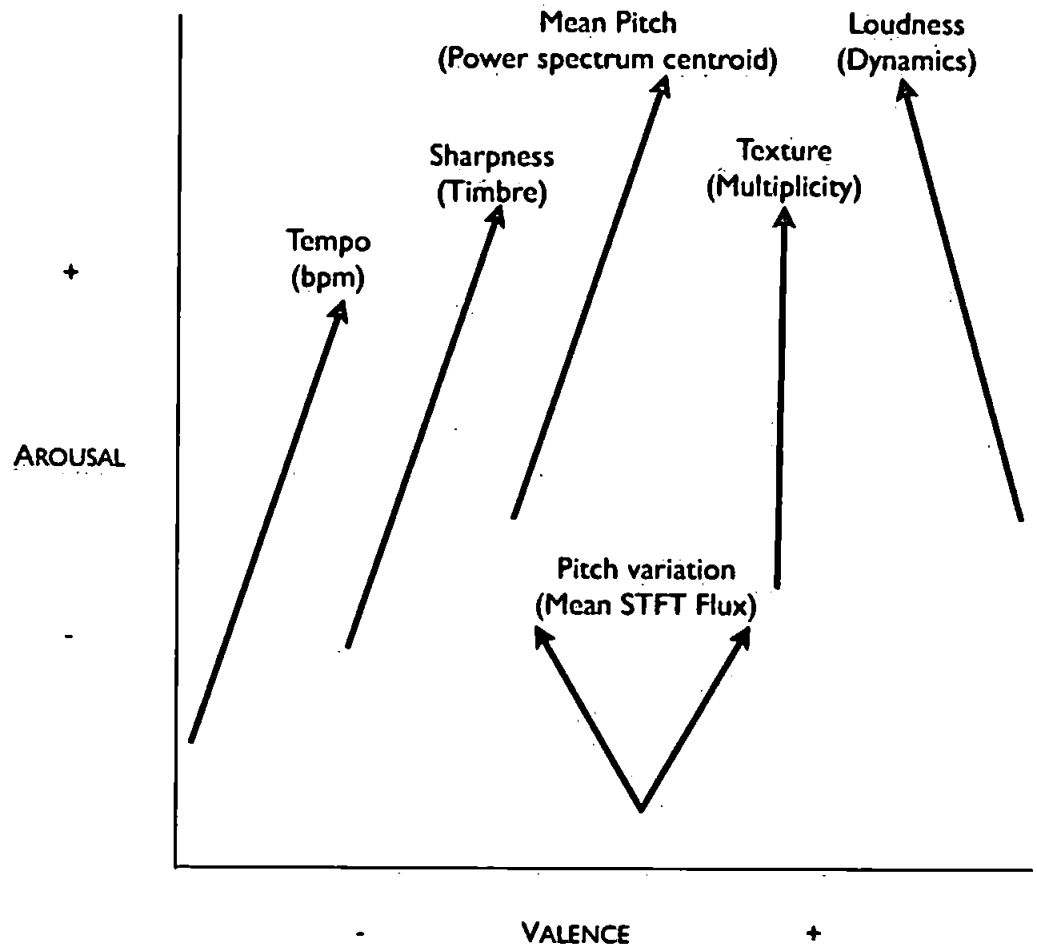


Figure 4.9: Qualitative representation of individual relationships between music variables and emotion appraisals: summary of observations from model analysis. The direction of the arrows indicates an increase in the variable indicated (the arrow sizes and angles formed with both axis are merely qualitative, and cannot be interpreted in mathematical terms)

Power Spectrum Centroid (mean pitch): the highest pitch passages relate to higher arousal and valence (quadrants 1, 2 and 4);

Sharpness (timbre): sharpness induced increased arousal and valence;

Mean STFT Flux (pitch variation): the average spectral variations related positively to arousal, for example large pitch changes are accompanied by increased activation. The pitch changes have both positive and negative

effects on valence, suggesting more complex interactions).

4.4 Discussion and Conclusions

Following an "emotivist" perspective (see Section 2.3.1) I considered that music can elicit affective experiences in the listener, and I focused on the sound features as a source of information about this process. Then, I presented and tested a novel methodology to study affective experience with music. By focusing on continuous measurements of emotion, the choice of the modelling technique considered two important aspects of sound: interactions between different features and their temporal behaviour. I proposed and tested recurrent neural networks as a possible solution due to their adaptiveness to deal with spatiotemporal patterns. To test the model Korhonen's experimental data (Korhonen, 2004a) was used.

The initial focus was on the reduction of psychoacoustic variables used by Korhonen, in order to identify the core group of variables relevant for this study hypothesis, but also to reduce the redundancy within the set. After preliminary simulations a set of 6 variables were selected, comprising dynamics (loudness), pitch level (spectral centroid), timbre (sharpness), pitch variation (mean spectral flux), texture (multiplicity) and tempo. Then a series of simulations were conducted to "tune" and test the model. To train the neural network to respond as closely as possible to the responses of human participants, 486 seconds of music (3 pieces) were used. An additional 632 seconds of music (3 pieces) served as a test data. The model did not know these three pieces. It was shown that the model predictions resembled those obtained from human participants. The generalisation performance validated the model and supported the hypothesis that sound features are good predictors of emotional experiences with music (at least for the affective dimensions considered).

In terms of modelling technique the model constitutes an advance in several respects. First, it incorporates all music variables together in a single model, which permits to consider interactions among sound features (overcoming some of the drawbacks from previous models Schubert, 1999a). Second, artificial neural networks, as non linear models, enlarge the complexity of the relationships between music structure and emotional response observed, since they can operate in higher dimensional spaces (not accessible to linear modelling techniques such as the ones used by Schubert, 1999a and Korhonen, 2004a). Third, the excellent generalisation performance (prediction of emotional responses for novel music stimuli) validated the model and supported the hypothesis that sound features are good predictors of the subjective feeling experience of emotion in music (at least for the affective dimensions considered). Fourth, another advantage, is the ability to analyse the model dynamics; an excellent source of information about the rules underlying input/output transformations. This is a limitation inherent in the previous models I wished to address. It is not only important to create a computational model that represents the studied process, but also to analyse the extent to which the relationships built-in are coherent with empirical research. In this chapter consistent relationships between music features and the emotional response, which support important empirical findings (e.g. Hevner, 1936, Gabrielsson & Juslin, 1996, Scherer & Oshinsky, 1977, Thayer, 1986, Davidson, Scherer, & Goldsmith, 2003; see Schubert, 1999a and Gabrielsson & Lindström, 2001 for a review), were found.

This study presented some evidence supporting the "emotivist" views on musical emotions. It was shown that a significant part of the listener's affective response can be predicted from the psychoacoustic properties of sound. It was also found that these sound features (to which Meyer referred as "secondary" or "statistical" parameters) encode a large part of the information that allows the approximation of human affective responses to music. Contrary to

Meyer's (Meyer, 1956) belief, the results presented here suggest that "primary" parameters (derived from the organisation of secondary parameters into higher order relationships with syntactic structure), do not seem to be a necessary condition for the process of emotion to arise (at least in some of its components). This is also coherent with Peretz et al. (1998) study, in which a patient lacking the cognitive capabilities to process the music structure (including Meyer's "primary" parameters), was able to identify the emotional tone of music.

Chapter 5

A psychophysiological study of musical emotions

In the previous chapter it was shown that judgements of affective responses to music can be modelled using only low level acoustic music features. The analysis of the model showed that loudness, tempo, pitch level and variations, sharpness and texture are good predictors of participants' emotional appraisals quantified as arousal and valence. Those results support the "emotivist" view on musical emotions. This chapter aims to verify experimentally those observations, and to extend the observed emotion components to physiological cues.

As discussed in Chapter 2, dimensional theories allow for a meagre representation of emotions. By asking participants to focus on their own feelings, self report of arousal and valence only controls for one component: the conscious feeling. In this study the improvement of this description is addressed by considering physiological cues. A new experiment was planned following a similar framework to the one presented in the previous chapter (Korhonen, 2004a; Schubert, 1999a), but with the additional measurement of physiological activity. The intention is to improve the description of the affective experience with music by accounting for other components of emotion. The goal is to assess

the relevance of physiological cues for the predictions of affective experiences with music. This chapter describes the psychophysiological framework prepared and the results obtained. Those results are to be used later (Chapter 6) to verify the model presented in this chapter, and to test the extension of it to physiological cues.

Modern theories of emotion usually consider physiological activation to be a basic component of emotion. As discussed in Chapter 2, the "component process model" views emotion as a construct of coordinated changes in physiological arousal, motor expression and subjective feeling, which may be highly synchronised to adapt in an optimal way to the eliciting circumstances. While "utilitarian" emotions (including basic emotions) lead to a more differentiated and proactive set of physiological and behavioural changes, aesthetic emotions (which differ in the appraisal concerning goal relevance) may produce more diffuse and reactive changes (Scherer, 2004).

Recent studies using physiological measurements have provided consistent evidence about the relation between affective states and bodily feelings (e.g. Harrer & Harrer, 1977; Khalfa et al., 2002; Krumhansl, 1997; Rickard, 2004). Krumhansl (1997) controlled for the widest spectrum of physiological variables (e.g. spectrum of cardiac, vascular, electrodermal, and respiratory functions) and some emotion quality ratings (e.g. happiness, sadness, fear, tension), reported by participants on a second by second basis. Krumhansl supports the idea that distinguishable physiological patterns are associated with different emotional judgements. In another series of studies (Witvliet & Vrana, 1996; Witvliet, Vrana, & Webb-Talmadge, 1998), among other variables, researchers investigated the effect of music on skin conductance and heart rate. Like others (e.g. Iwanaga and Tsukamoto (1997); Khalfa et al. (2002); Rickard (2004)), these studies have shown that heart rate (HR) and skin conductance response (SCR) increase with arousing or emotionally powerful music.

Although evidence of an emotion specific physiology was never found (Ekman & Davidson, 1994; Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 1993), research in peripheral feedback provides evidence that body states can influence the emotional experience with music (Dibben, 2004; Philippot et al., 2002). Research on emotion has delivered strong evidence that certain patterns of physiological activation are reliable references of the emotional experience. Peripheral feedback has also been considered to be able to change the strength of an emotion even after this has been generated in the brain (Damasio, 1994). Physiological arousal has also been associated with psychological representations and determinants of emotion, such as valence (or hedonic value) and arousal (Lang et al., 1998).

5.1 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. Participants were asked to report the emotion "felt" while listening to the music (rather than "thought" to be expressed by the music (Gabrielsson, 2002)). Participants' heart rate (measured in bpm) and skin conductance level (SCL) (measured in μS) activity was also recorded.

This experiment permits to investigate the existence of relevant interactions between the psychological and physiological concomitants of emotional experience with music.

The hypotheses for this new experiment are the following:

1. Tempo, loudness, sharpness, texture, pitch level and pitch variation are expected to have significant relationships with changes in the subjective feelings component;

2. Music can alter both the physiological component and the subjective feeling

of emotion in response to music, sometimes in highly synchronised ways.

5.2 Method

5.2.1 Participants

Forty-five volunteers participated in the experiment (Appendix A contains the call for participants used to publicise 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). At the beginning of the experiment, participants answered a questionnaire regarding their musical education and experience, exposure to and enjoyment of classical music. The results are shown in Figure 5.1. Participants used 5-point Likert scales to answer these questions (the questionnaire is included in Appendix B). 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 classical music. Participants also reported a high level of enjoyment of this music style (the mean rating was 4.2).

5.2.2 Equipment

Each participant sat comfortably in a chair inside a quiet room, and listened to the music via Sennheiser closed headphones (through a M-Audio sound interface connected to a computer). The physiological measures were obtained using

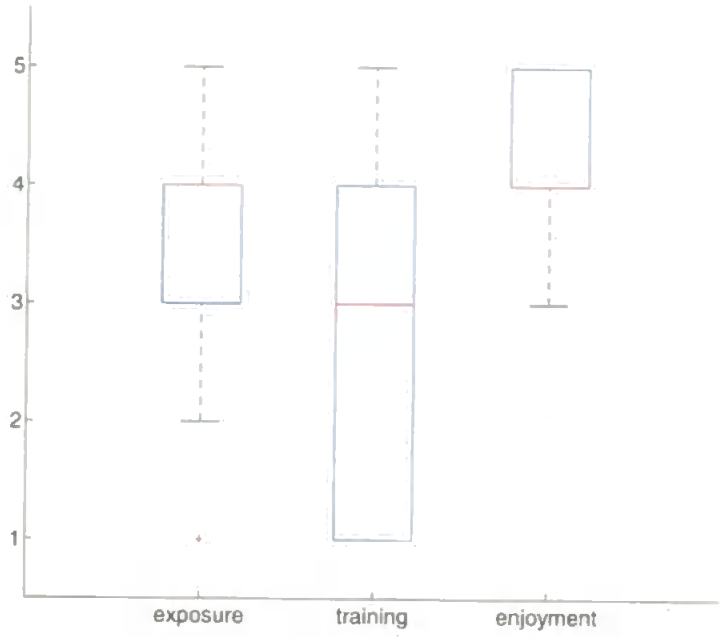


Figure 5.1: Experiment questionnaires: The participants' answers to each question are depicted by the smallest observation, lower quartile, median, upper quartile, and largest observations.

a WaveRider biofeedback system (MindPeak, USA). Leads were attached to the subjects for measuring heart rate and skin conductance level. Participants reported their emotional state by using the EMuJoy software (Nagel, Kopiez, Grewe, & Altenmüller, 2007), a computer representation of a two-dimensional emotional space (2DES). The self report data was later synchronised with physiological data.

5.2.3 Stimuli

The stimulus materials consisted of 9 pieces chosen by two professional musicians (one composer and one performer, other than the author), to illustrate the combination of valence and arousal, and to cover the widest area of the 2DES possible (combinations of arousal and valence factor). 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 5.1, and described below.

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

Table 5.1: Pieces used in the experiment.

The following paragraphs contain a description of the music pieces used in the experiment. They describe the general character of the pieces and are not necessarily linked to its emotional character. For each piece, the expected emotional character, which is on the basis of their selection, is explicitly indicated

(at the end of each description) in terms of arousal and valence factors.

Piece 1 Adagio is a piece for strings and organ[1], in the key of G minor, that leads to a mood of solemnity, with occasional outbursts of melancholy (and tragedy). This piece is expected to be belong to quadrant 3 (low arousal, negative valence).

Piece 2 "In the Hall of the Mountain King" is a piece of orchestral music composed by Edvard Grieg for "Peer Gynt", a theatre play by Henrik Ibsen. The piece represents Peer Gynt's (the adventurer character) attempts to escape from the King's castle, after sneaking in and insulting his daughter. The theme begins slowly and quietly evolving through low registers, where the bassoons represent Peer's careful and quiet movements. Then, the theme is smoothly modified with the appearance of different instruments playing sets of ascending notes (representing the appearance of the King's trolls). Subsequently the tempo gradually speeds up (the trolls spot Peer) and the music becomes increasingly louder, until a series of crashing cymbals and timpani rolls take over, and the piece concludes (Peer escapes successfully). This piece is expected to elicit responses along the 1^{st} quadrant of the 2DES (positive arousal and positive valence).

Piece 3 Piece 3 is a prelude in G-major from the 1^{st} 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 1^{st} quadrant of the 2DES, with higher arousal during the second part.

[1]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.

Piece 4 Beethoven's "Romance No. 2" in F major is a piece in the style of a concerto movement, although less complex [2]. Noted as an "adagio cantabile", it presents melodies with lyrical characteristics, characterized by the interplay between the soloist and the orchestra. This piece is called romance for its light, sweet tone, almost like a song. This piece is expected to induce low to high levels of arousal and positive valence (quadrants 1 and 4).

Piece 5 Chopin's Nocturne no 2 (E-flat Major), published in 1833, is a piece with a romantic character, flowing with an expressive and dreamy melody, although with a certain melancholic style. Like the majority of Chopin's nocturnes it is built upon a simple A-B-A form. This piece is expected to elicit low arousal and positive valence (quadrant 4).

Piece 6 The second movement of Mozart's "Divertimento" in B-flat major (written in Salzburg in early 1772) is a piece with a dynamic ("loose") structure and a relaxed feeling. It 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 1).

Piece 7 "Jeux de Vagues" (Frolics of waves) is the 2^{nd} movement of Debussy's "La Mer". This piece, in C sharp minor, 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's a piece of variety and "colour" 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; low arousal and negative valence is not an expected result).

[2] It might have been written as a slow movement to an earlier, unfinished Violin Concerto, written between 1798 and 1802

Piece 8 Liebesträume ("Dreams of Love") No. 3 in A Flat Major, is the last of 3 solo piano works published by Liszt in 1850. The 3 works were composed after the poems by Ludwig Uhland and Ferdinand Freiligrath, which depict three different forms of Love. The poem for the third "notturno" describes mature love[3]: "Love as long as you can! The hour will come when you will stand at the grave and mourn". The piece is usually divided into three sections separated by a fast "cadenza", with the same melody maintained during the entire piece, however each time varied. This piece is expected to elicit responses of low arousal (quadrants 3 and 4).

Piece 9 The "Ciaccona" ("Chaconne") is the concluding movement of Bach's "Partita no. 2" that lasts some 13 to 15 minutes. This piece is considered to be a pinnacle of the solo violin repertoire, but several different transcriptions of the piece have been made for other instruments. In this experiment a transcription for piano by Ferruccio Busoni (performed by Mikhail Pletnev) was used. This piece is in D minor, with the middle section in major mode, and simple triple time. This piece is expected to elicit responses in the 4 quadrants.

5.2.4 Psychoacoustic encoding

The sound features that quantify the music stimuli were obtained using PsySound 3 (Cabrera, 1999), BeatRoot (Dixon, 2001) and the melodic contour extractor by (Dittmar, Dressler, & Rosenbauer, 2007). The six key variables used in Chapter 4 (loudness, tempo, power spectrum centroid, multiplicity, spectral flux and sharpness) were calculated again. Three more variables were also estimated: melodic pitch, timbral width and roughness. The 9 psychoacoustic variables chosen are shown in Table 5.2.

Loudness, tempo, power spectrum centroid, multiplicity, mean STFT flux and

[3]The first work describes exalted love while the second depicts erotic love.

Sound feature	Group	Alias
Dynamic Loudness	Loudness	L
Beats-per-minute (bpm)	Tempo	T
Power Spectrum Centroid	Mean Pitch	P
Multiplicity	Texture	Tx
Mean STFT Flux	Pitch Variation	Pv
Melodic Pitch	Melody Contour	mP
Sharpness	Timbre	S
Timbral Width	Timbre	TW
Roughness	Roughness	R

Table 5.2: Psychoacoustic variables considered for this study. For convenience the input variables will be referred to with the aliases indicated in the table throughout this chapter.

sharpness (the six key variables used in the previous chapter) were already described in Chapter 4 (Section 4.1.2). Timbral width was also described in that chapter and, although eliminated from the inputs set for the model presented in the same chapter (the aim was to reduce the number of features in each psychoacoustic group to one - sharpness was the chosen feature), it is tested again as an attempt to improve the description of timbre (which is multidimensional in nature). Timbral width, referred to as Ti_2 in Chapter 4, was preferred to Mean STFT Rolloff (Ti_3) and to Standard Deviation STFT Rolloff (Ti_3), because it has shown the second best performance ($rms_{Ti_2} = 0.089$) and the valence predictions for novel music ($rms_{Ti_{2-testV}} = 0.080$) when compared with the remaining features ($rms_{Ti_{3-testV}} = 0.121$, $rms_{Ti_{4-testV}} = 0.087$). The melodic pitch (mP), estimates the contour of the main melodic line of polyphonic sounds, and it was included to have a description of the main melody (predominant voice) contour (which in the previous chapter was approximated using the Mean STFT Flux). Finally, Roughness is a basic psychoacoustical sensation for rapid amplitude variations which reduces the sensory pleasantness and the quality of noises. The sound features extracted with PsySound were down-sampled from the original sample rates to 1Hz in order to obtain second by second values. A

Figure 5.2: Experimental framework: participants listened to the music using a pair of closed headphones. In front of them, a computer screen shows the EMuJoy interface for the self report of Arousal and Valence. Leads were attached to measure Heart Rate and Skin Conductance Response.

visualisation of each feature for each piece of music used in the experiments is displayed in Appendix F.

5.2.5 Procedure

The goal of the experiment was explained in a standard way to each participant. The quantification of emotion was addressed and the self report framework was introduced. Then, the equipment for physiological measurements was prepared and leads were attached to participants in order to measure their heart rate (HR) and skin conductance level (SCL). Participants were given written instructions (see "Participants Information Sheet" in Appendix C). Figure 5.2 shows the experimental set up.

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 (IAPS) manual (Lang, Bradley, & Cuthbert, 2005) was selected (see Appendix D), in order to represent emotions in the 4 quadrants of the 2DES (2 per quadrant), as well as the neutral affective state (centre of the axis). The pictures were shown in a nonrandomised order, in order to avoid starting or finishing the picture slideshow with a scene of violence. Each picture was shown for 30 seconds each, with a 10s delay inbetween presentations. The only aim of this exercise was to get participants comfortable 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 expressed by the music. When ready, the experiment started and the first piece was played. The pieces were presented in a randomised order (see Appendix E), with a small break of 15 seconds between each piece (unless the participant needed more time).

An experiment 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 acclimatise and settle into the location.

5.2.6 Experiment design

In this experiment there are four dependent variables: the self report of subjective feeling (arousal and valence) and the physiological responses (heart rate and skin conductance level). The independent variables are the nine music pieces.

5.3 Results

The analysis of the observed variables is divided into three steps. First, the mean values across all participants were calculated on a second by second basis for each piece. Then, the expected subjective feeling responses for the chosen pieces are compared to the experimental data. Subsequently, the pieces are segmented and the relationships between the mean levels of each experimental variable are investigated. In the last step, each pieces is analysed on a second by second basis.

5.3.1 Data processing methods

The arousal and valence for each participant was recorded overtime from the mouse movements. These values were normalised to the continuous scale raging from -1 to 1, with 0 as neutral. Then, arousal and valence were averaged across all participants, on a second by second condition, for each music piece. No further processing was done on this data.

For each participant the physiological variables had to be processed to rule out the effects of individual differences on physiological levels, because the physiological activity levels are highly dependent on individual characteristics. Two possibilities were assessed: the calculation of relative and absolute values. While the first consists on dividing the physiological activity of each participant by its individual baseline readings obtained in a nonstimulus condition, the second consists on subtracting that quantity. The first method was chosen for two reasons: first because it involves fewer calculations and thus is less prone to error; second because it is more intuitive. By calculating the relative values of physiological levels, the quantity obtained is automatically comparable between subjects, since it consists of the relative deviations from participants individual baselines which is represented as 1.0. Using the subtraction method those

values still need to be normalised in order to be comparable, involving further calculation, since the baselines will be different for each participant. Using the division method, any deviation in the physiological variables is also more intuitive and easily interpreted either as an increase (higher than 1.0) or decrease (lower than 1.0) relative to the baseline. For these reasons the individual heart rate and skin conductance level readings were divided by the average of the 20 seconds individual baseline readings obtained in a nonstimulus condition before the experiment started. The resultant time series correspond to the relative values for hear rate (rHR) and skin conductance level (rSCL).

Another operation was performed on the SCL to obtain values reflecting the skin conductance response. Since the recordings of SCL obtained yield the absolute values of individual skin conductance, they depend on several factors such as the moisture level of the skin, temperature and blood flow. As a standard procedure, by differentiating SCL, we obtain the SCR, which reflects variations in SCL, and so quantifying changes, rather than the general level of perspiration. To do so, the rSCL values were differentiated (1^{st} order differentiation) in order to obtain rSCR, the standardised values of the skin conductance response. The final time series are shown on the bottom of Figures 5.3 to 5.11.

The analysis procedure is divided in three steps: the analysis of whole music pieces, the analysis of pieces segments and the analysis of continuous measurements.

The first analysis (whole music pieces) aims at verifying the expected emotional value of each piece, by comparing the predicted quadrants of the 2DES covered by each piece with the experimental results. The relationships between the "knowing" and "liking" each piece (variables obtained from the participants questionnaires) are also investigated in order to assess any influence of these factors on participants responses.

In the second step the pieces are segmented in order to observe and analyse

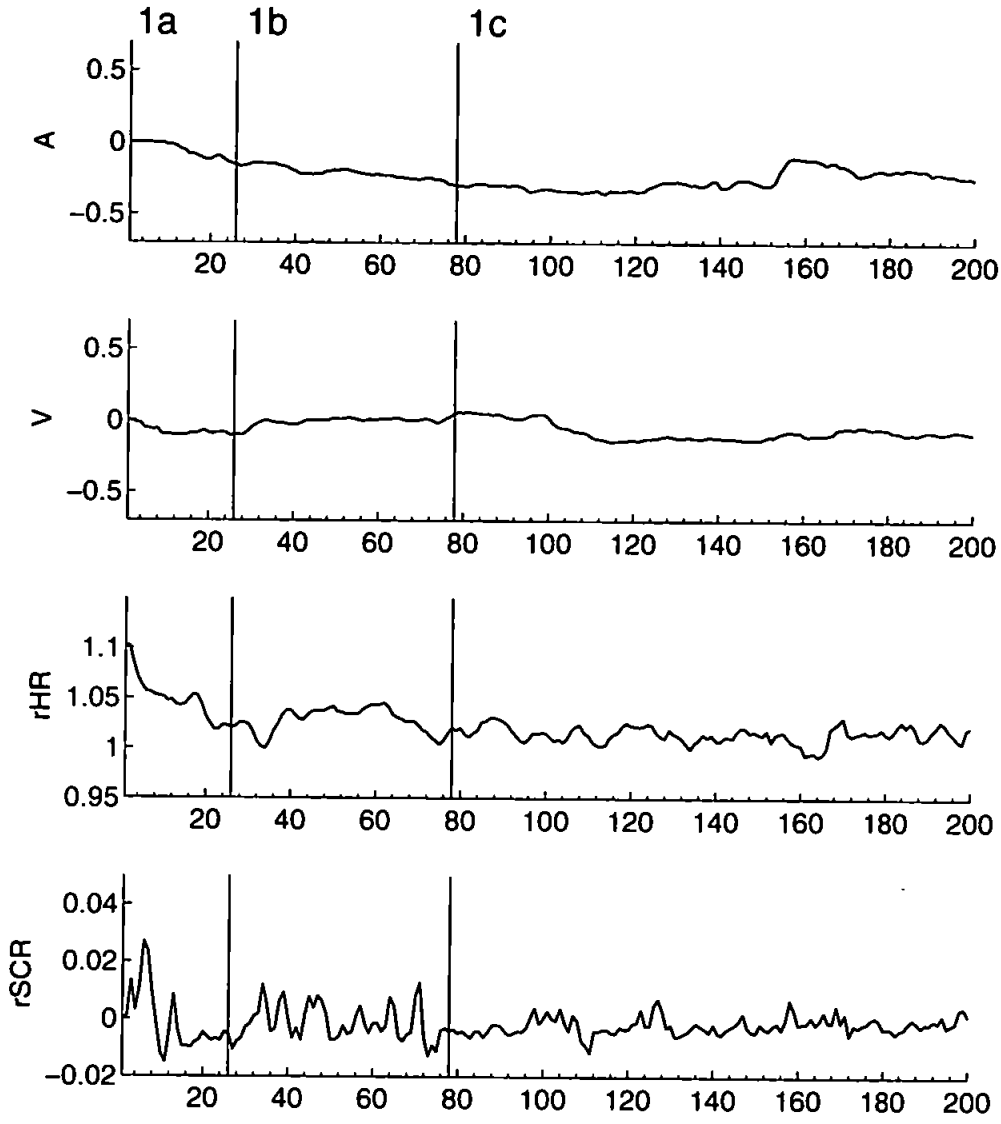


Figure 5.3: Albinoni, *Adagio* (Piece 1): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants.

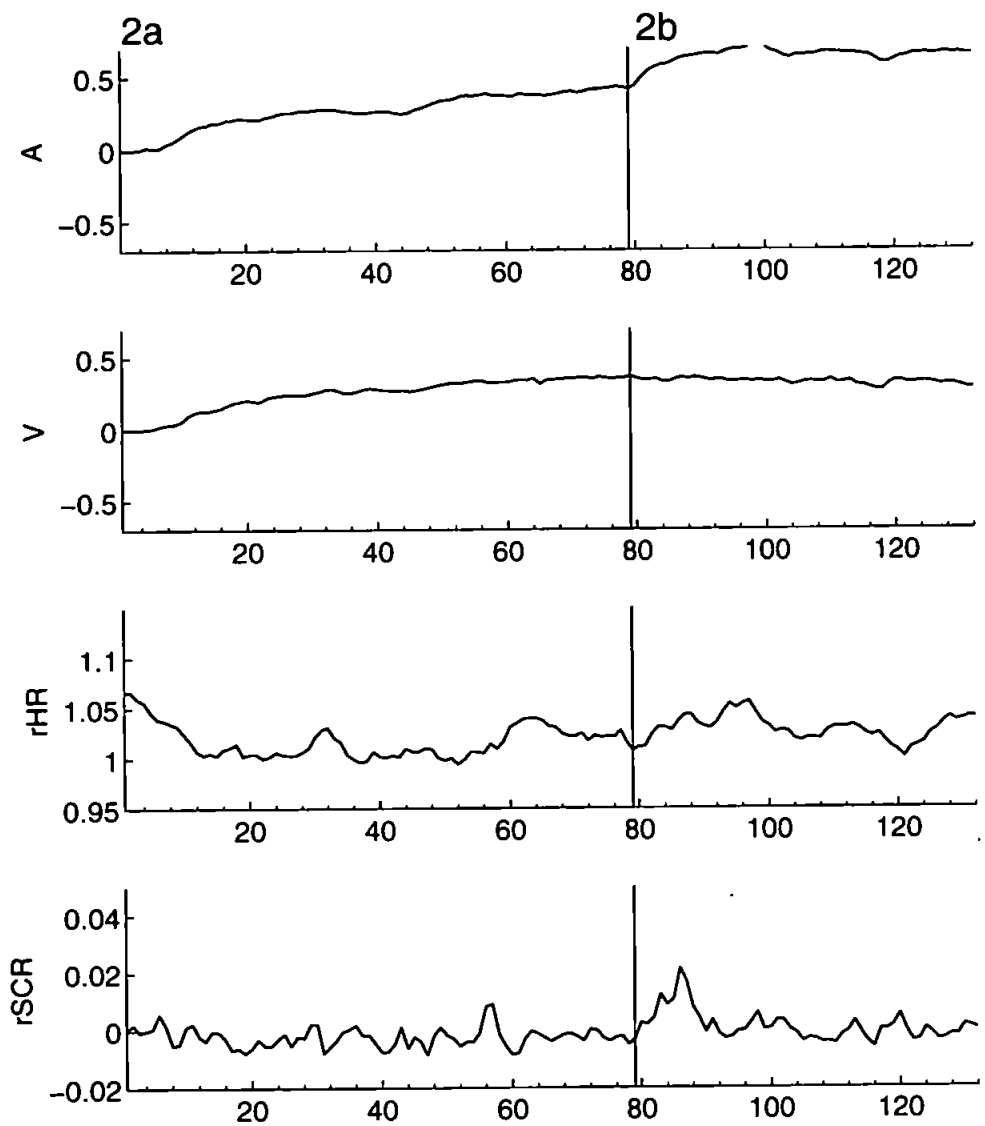


Figure 5.4: Grieg, *Peer Gynt Suite No. 1* (Piece 2): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants.

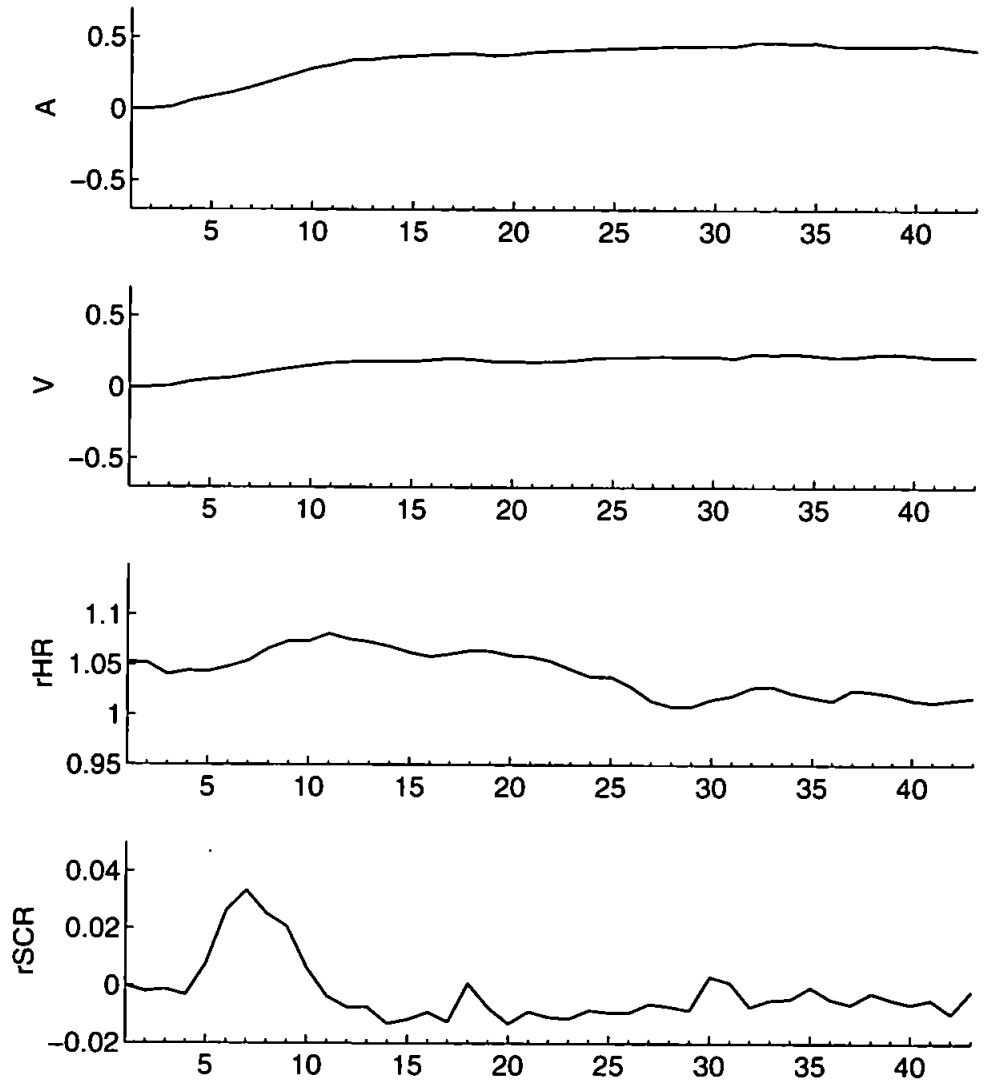


Figure 5.5: Bach, *Prelude and Fugue No. 15* (Piece 3) : mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants.

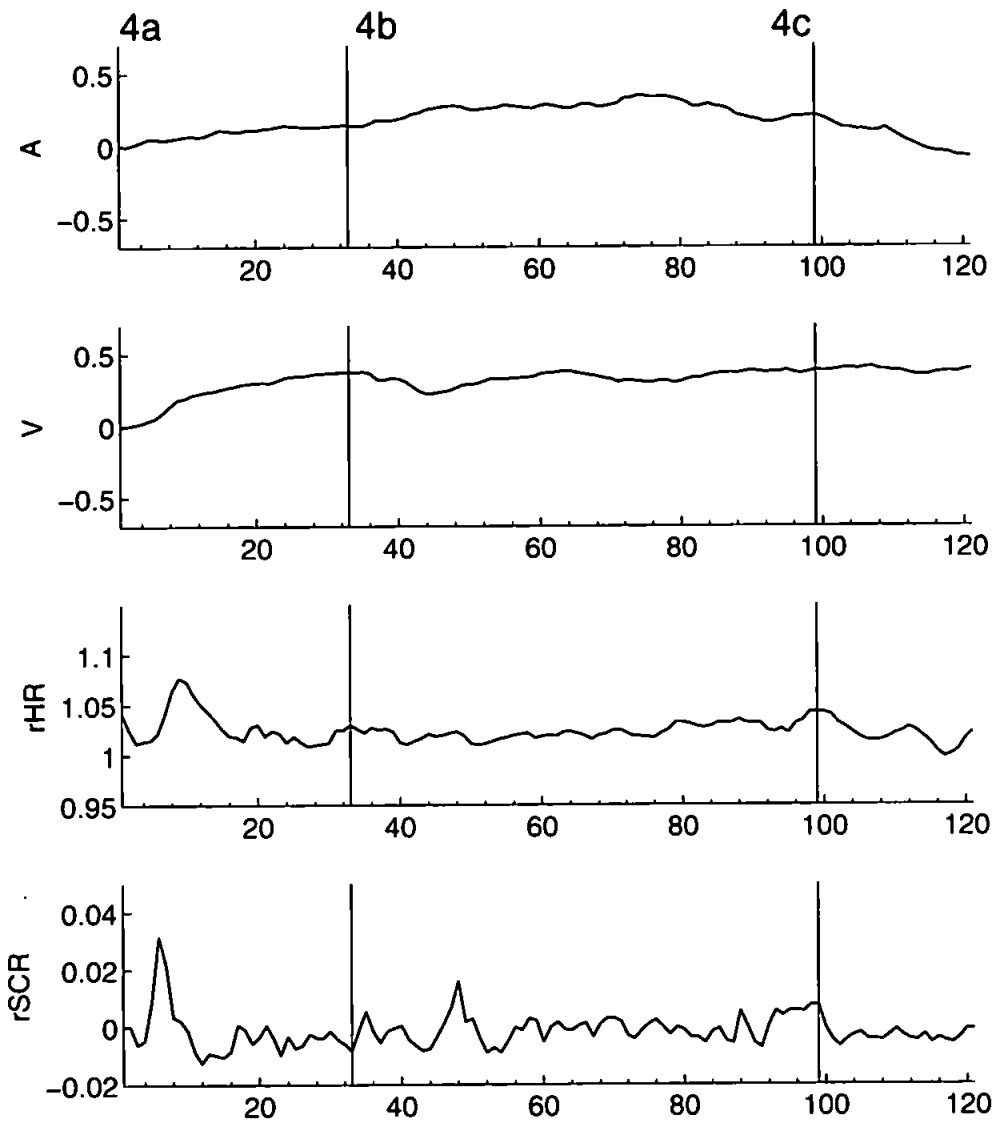


Figure 5.6: Beethoven, *Romance No. 2* (Piece 4): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants.

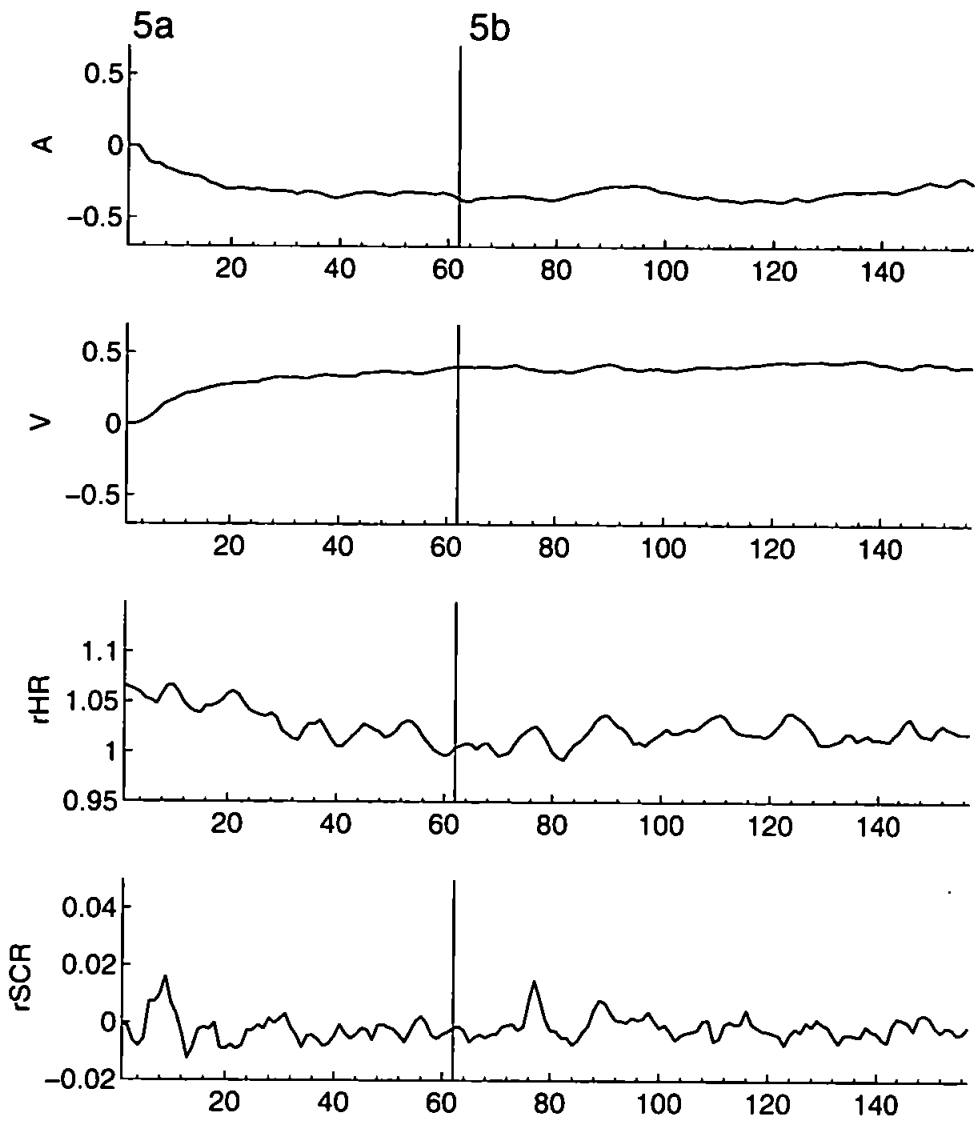


Figure 5.7: Chopin, *Nocturne No. 2* (Piece 5): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants.

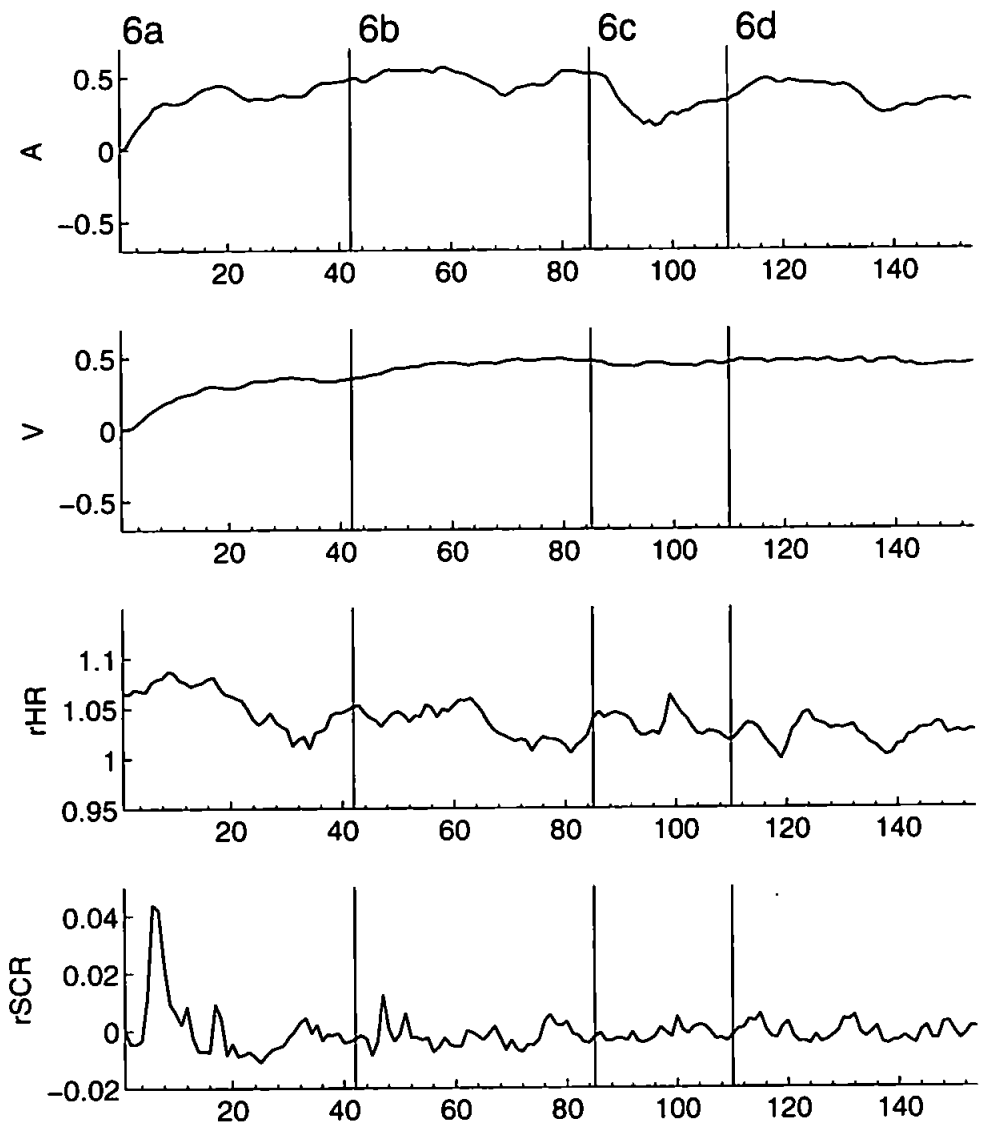


Figure 5.8: Mozart, *Divertimento* (Piece 6): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants.

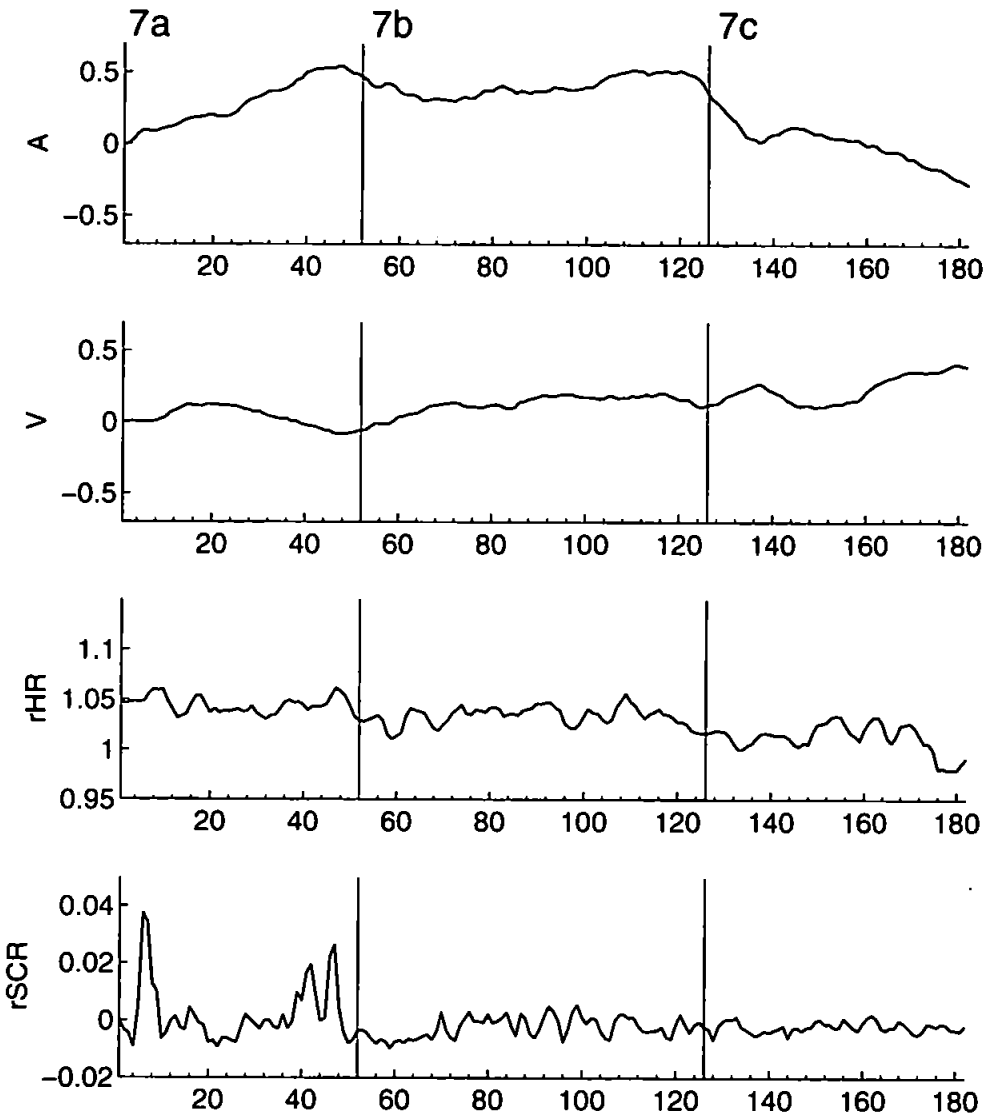


Figure 5.9: Debussy, *La Mer* (Piece 7): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants.

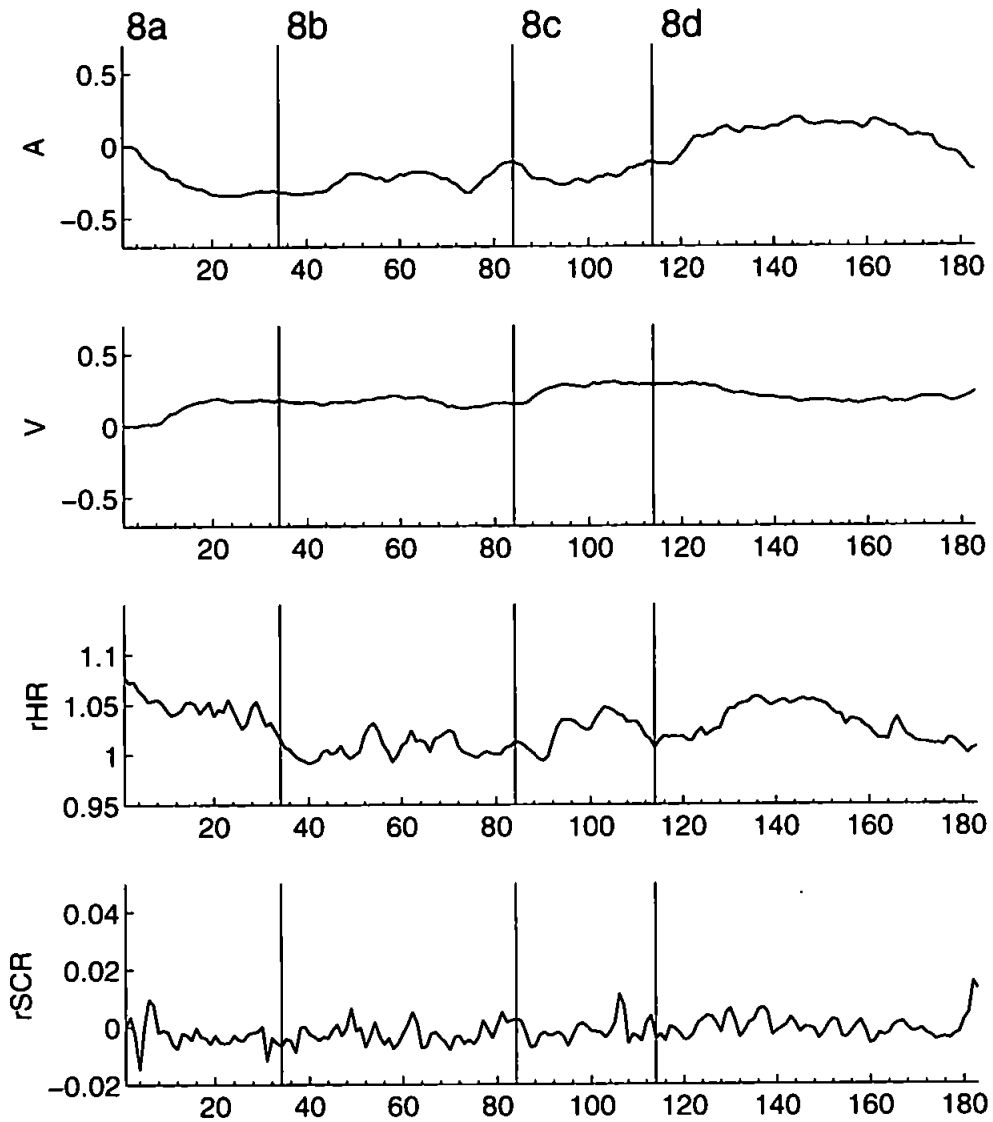


Figure 5.10: Liszt, *Liebestraum No.3* (Piece 8): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants.

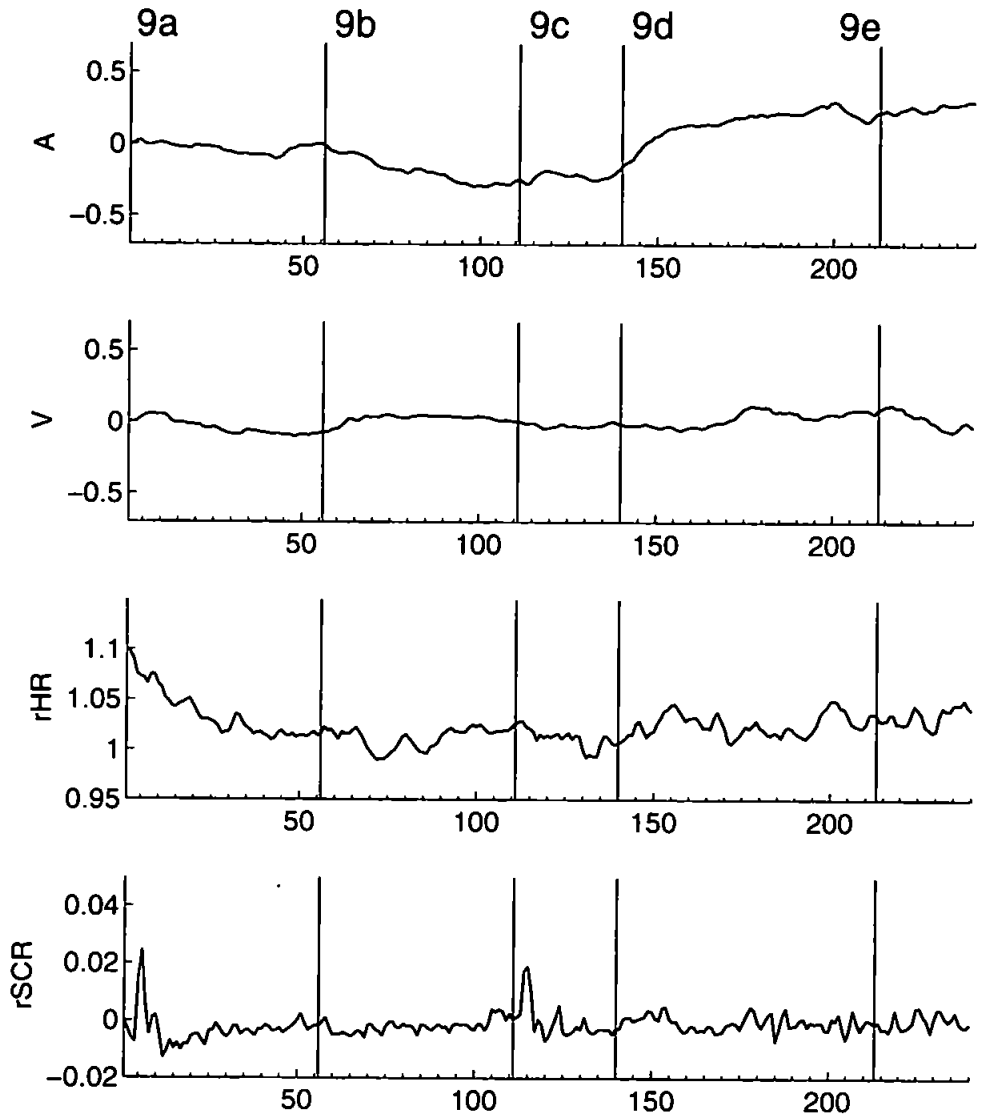


Figure 5.11: Bach, *Partita No. 2* (Piece 9): mean Arousal, Valence, Heart Rate and Skin Conductance Level over all participants.

the different patterns within each piece in more detail. This procedure aims to separating the pieces accordingly to distinct affective responses and to obtain a set of observations that allow for statiscally valid analysis. The correlations between the mean levels of the sound features and the arousal/valence pair, and also between sound features and physiological activity, are assessed in order to investigate the relationships between the mean levels of each variable (e.g. how the segments with faster tempi are correlated to the mean levels of physiological activity for the same segment).

Finally, in the third step, the relationships between physiological activity and self report are analysed in more detail, since the physiological variables are not expected to show simple relationships with the self report variables. The previous analyses, by considering the mean levels for each piece or segment, do not account for the temporal relationships between variables. This analysis focuses on the synchronisation between psychological and physiological events, in order to assess the relevance of the peripheral feedback hypothesis.

5.3.2 Analysis of whole music pieces

Table 5.3 shows the experimental data mean averages and standard deviations, per piece, across all participants, whose location in the A/V space is shown in Fig. 5.12a). The same concept was used to represent the pairs rHR/drSCL (5.12b). The pieces were grouped according to the A/V quadrant that they belong to (quadrants 1 to 4 anti-clockwise; 1 and 2 positive A, 1 and 4 positive V).

The majority of the pieces (2, 3, 4, 6, 7 and 9) belong to the 1^{st} quadrant (A>0, V>0). Only M1 belongs to quadrant 3 (A<0, V<0), while M5 and M8 belong to quadrant 4 (A<0, V>0). There are no pieces with average report on quadrant 2. Nevertheless, because some of the pieces show a high variability (see Table 5.3) in arousal and valence, they cover more than one quadrant (especially pieces 7 and 9). That is shown in Figure 5.13, where the second by second values of

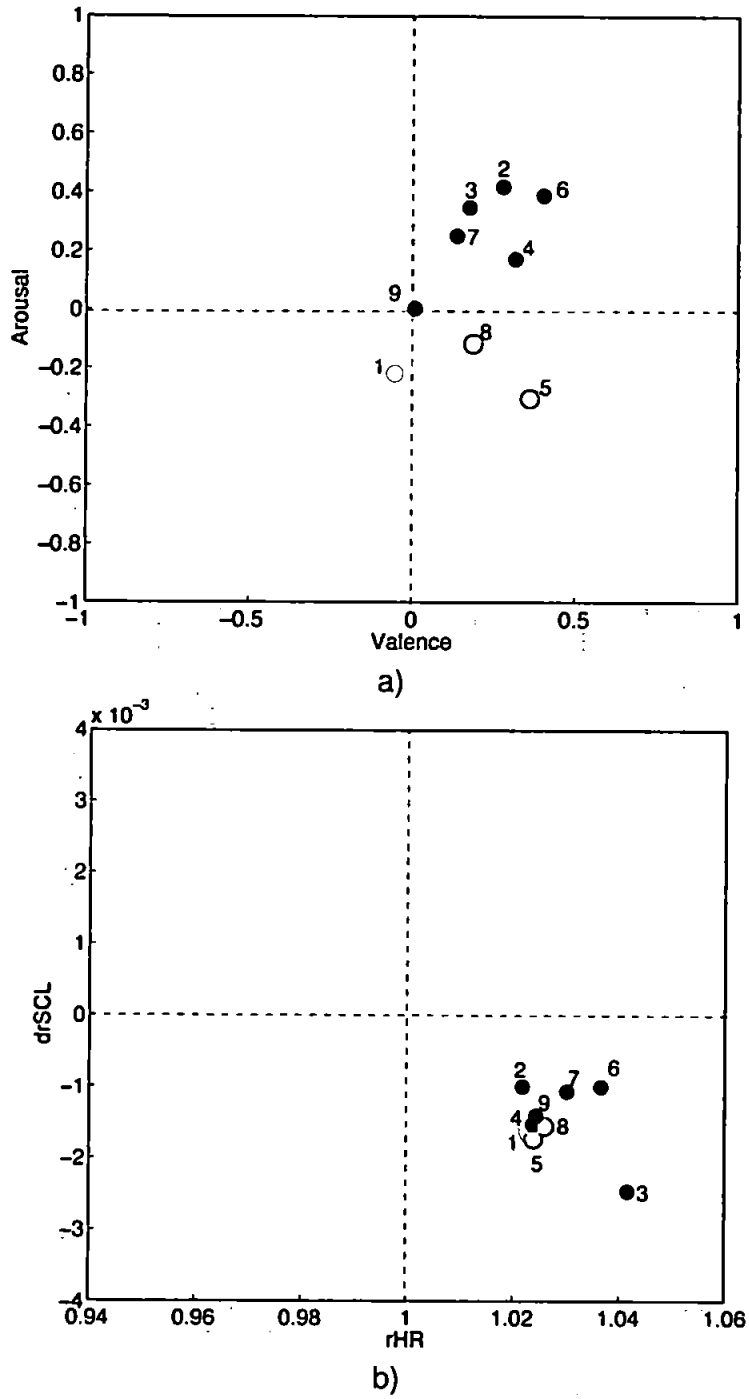


Figure 5.12: Mean Arousal and Valence (a) and diff. mean Skin Conductance Level and Heart Rate (b) for each piece. White and black dots correspond to the pieces with positive Valence, and black and dark grey dots indicate the pieces with positive Arousal. The numbers indicated next to each dot correspond to the piece ID, as indicated in Table 5.3.

Piece ID	A		V		rHR		rSCR	
	mean	std	mean	std	mean	std	mean	std
1	-0.220	0.089	-0.054	0.057	1.023	0.017	-0.0016	0.005
2	0.421	0.208	0.277	0.092	1.022	0.016	-0.0010	0.005
3	0.347	0.142	0.174	0.068	1.042	0.022	-0.0025	0.011
4	0.166	0.112	0.316	0.087	1.024	0.012	-0.0014	0.006
5	-0.305	0.071	0.361	0.094	1.034	0.016	-0.0017	0.004
6	0.387	0.113	0.402	0.102	1.036	0.020	-0.0010	0.007
7	0.243	0.217	0.138	0.114	1.030	0.017	-0.0011	0.006
8	-0.118	0.170	0.187	0.065	1.026	0.020	-0.0016	0.004
9	0.001	0.188	0.007	0.054	1.024	0.018	-0.0014	0.004

Table 5.3: Mean and standard deviation values for arousal, valence, heart rate and galvanic skin response across all participants for each pieces.

arousal and valence were plotted in the 2DES.

The experimental results showed a low range of values, with a tendency for the first quadrant of the 2DES. There are very few moments of negative valence. Although the results cover most parts of the expected areas of the 2DES they are concentrated in a small area. A possible reason for this is that the pieces lack in stimuli with such affective qualities. Another might be related to a general positive effect of music on mood.

No significative correlations were found between participants knowing or liking each piece ($r = 0.372$ - mean for all pieces). Nevertheless the two pieces that more participants knew (Piece 2 and Piece 5) had also the highest ratings of enjoyment. No significant correlations were found between the mean values of experimental data and the questionnaire variables, but valence tends to be significantly higher for pieces that participants liked more (r=0.620, p=0.075).

5.3.3 Analysis of music segments

In order to observe and analyse the different patterns within each piece in more detail, the pieces were segmented. This procedure aims to separate the pieces accordingly to distinct affective responses and to obtain a set of observations

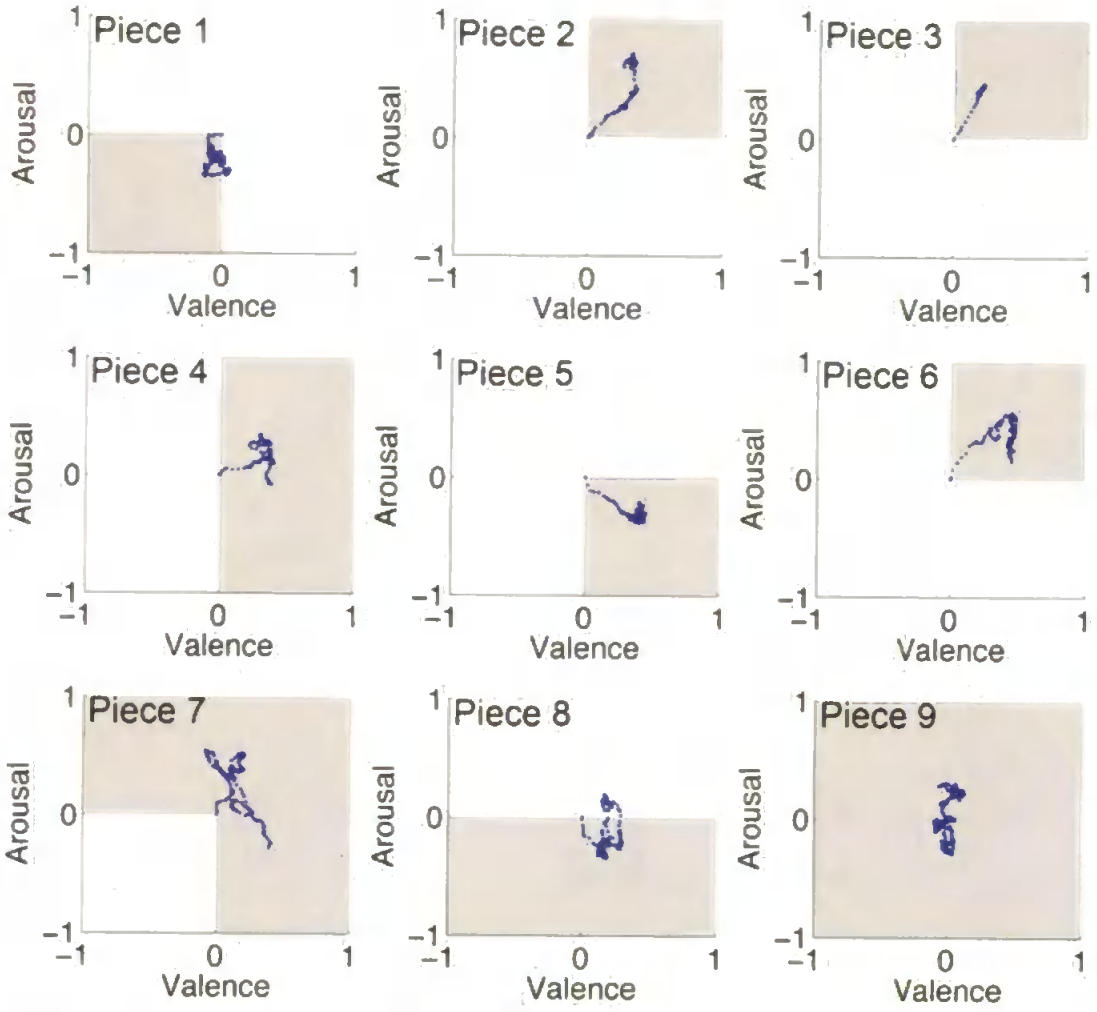


Figure 5.13: The figure shows the second by second values of Arousal and Valence, averaged across participants, for each piece used in the experiment. The grey rectangles indicate the areas of the 2DES, which correspond to the subjective feelings of emotion expected to be elicited in the listeners by each piece.

which allows for a statiscally valid correlation analysis. A professional composer was asked to divide each piece into different segments by focusing on criteria related to its form and "sense" of structure, and perceived affective value. If the segments selected synchronised with significant changes in self report for that segment, they were used as segmentation points.

The segmentation of each piece is shown in Table 5.4. The segments are also shown with the time series plots in Figures 5.3 to 5.11. Throughout this chapter each segment will be identified by their 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: 1b). Figure 5.14 shows a representation of the mean psychological and physiological activity for each segment.

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 5.4: Number of segments and correspondent sound file times for each piece.

Two separate tests were conducted. The first assesses the relationships between sound features and the arousal/valence pair. The second evaluates the correlations between sound features and physiological activity. The Spearman's rank correlation coefficient (rho) was used as a non-parametric measure of correlation between variables.

The initial ten seconds of the first segment of each piece were not included,

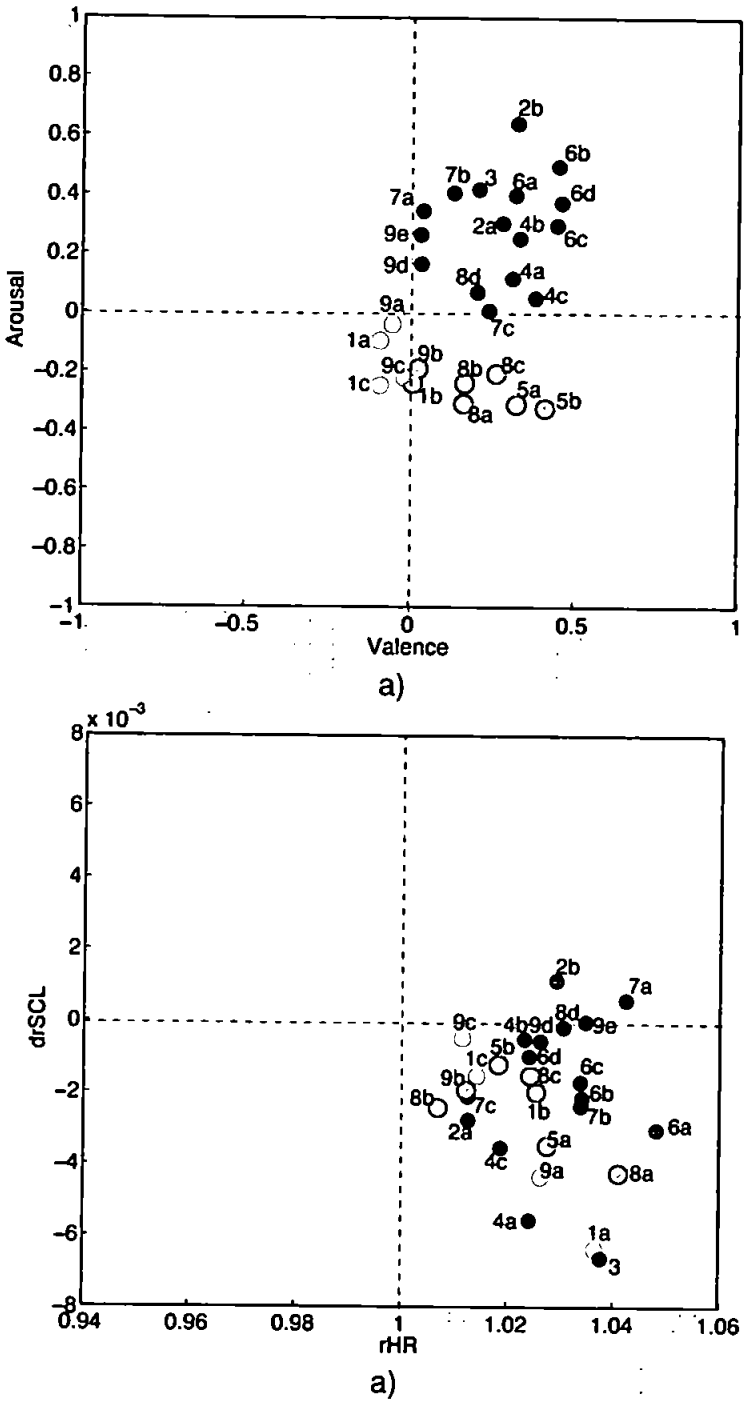


Figure 5.14: Mean Arousal and Valence (a), and Skin Conductance Response and Heart Rate (b) of the 27 segments of music. White and black dots correspond to the pieces with positive Valence, and black and dark grey dots indicate the pieces with positive Arousal. The label indicated next to each dot correspond to the segments identifiers as indicated in Table 5.4.

because the changes in physiological activity (especially SCR) during the initial period of each piece can be explained as an orientation response to the new stimulus. The self report during this period is also considered to be a moment of transition that allows participants to move the mouse to the desired position in the 2DES.

Sound features vs. physiological activity There is only one statistically significant correlation between the mean levels of sound features and physiological activity for each segment (see Table 5.5). It relates the level of HR in the segments with tempo ($rho(27) = 0.425, p =< 0.05$).

variables	rHR	rSCR
L	0.044	0.300
T	0.423*	-0.087
P	-0.032	0.007
C	0.147	-0.107
mP	0.060	-0.115
Tx	-0.068	-0.280
R	0.111	0.228
S	0.084	0.000
TW	0.123	0.123

Table 5.5: Correlations between sound features (L, T, P, C, mP, Tx, R, S and TW) and physiological activity (rHR - relative value of the Heart Rate level; rSCR - relative value of the Skin Conductance Response). * $p < 0.05$

Sound features vs. self report variables Sound features and self report variables have several significant correlations (see Table 5.6).

The strongest linear correlations with arousal in the segments considered are with L, T, P, S, and TW. These variables show positive relationships with the level of arousal in the segments. Valence correlated significantly with tempo ($rho(27) = 0.469, p =< 0.05$) and mean pitch ($rho(27) = 0.385, p =< 0.05$).

These results are coherent with the analysis of the model presented in Chapter 4, which have shown the relevance of these features to the prediction

variables	Arousal	Valence
L	0.571**	0.056
T	0.579**	0.480*
P	0.606**	0.385*
C	-0.130	-0.141
mP	-0.162	0.134
Tx	-0.209	0.187
R	0.062	0.074
S	0.634**	0.252
TW	0.589**	-0.095

Table 5.6: Correlations between sound features (L, T, P, C, mP, Tx, R, S and TW) and physiological activity (Arousal and Valence). ** $p < 0.01$, * $p < 0.05$

of self report of emotions expressed by the music. Timbral width, a variable not included in the previous chapter and tested here, is also significantly correlated to arousal and valence.

Physiological activity vs. self report variables No significant correlations were found between sound features and the physiological activity mean levels.

5.3.4 Analysis of continuous measurements

The previous analysis focused on the mean levels of music sound features, psychological and physiological activity, and in their relationships. As expected, the majority of significant correlations found showed relationships between sound features and self report data. Nevertheless, tempo was also found to be associated with heart rate level. In this section, the relationships between physiological activity and self report are assessed by investigating events synchronizations and by analysing the time series of each variable. Moreover, to assess the peripheral feedback hypothesis, is it relevant to identify the relationships physiological and psychological events and specifically the synchronicity (or close temporal relatedness) between them.

For the analysis of second by second interactions between physiological

dynamics and self report changes I focused on a methodology suitable to identify the strongest changes in the four experimental variables. Using a procedure similar to the one used by Grewe et al. (2005), a criterion was established in order to identify affective events among the self report and physiological time series. Considering the moments of strong changes in self report as indicators of such an event (moments when participants move their mouse and re-evaluate their emotional state) the differences between consecutive time steps in each of the time series were calculated.

The 95^{th} percentile of the absolute value of these changes and of the average SCR values[4], were established as the criteria to identify them. The 10s initial period of each piece was excluded from the analysis of strong changes. The changes in physiological activity (especially SCR) can be explained as an orientation response to the new stimulus (Grewe et al., 2005). The changes in self report in the same period were also excluded because they correspond to the transition from the neutral position to participants' individual rating.

Each value in each time series over the average threshold of all pieces was labelled as a "peak" (or strong change). The values shown in Table 5.7 are the sum of the total number of peaks in arousal, valence, heart rate and skin conductance response, for each music piece.

Piece 5, the piece with the lowest mean arousal of the experiment, had only one strong change in valence. The highest number of strong changes occurred in the pieces covering a wider area of the 2DES (Pieces 6, 7, 8 and 9). The correlations between the number of peaks in self report and strong physiological changes are shown in Table 5.8. Only the number of peaks in arousal correlated significantly to strong changes in heart rate ($r(9) = .909, p =< .001$). Although skin conductance response shows high linear correlation coefficients with arousal ($r(9) = .558, p == .119$) and heart rate ($r(9) = .580, p =< .102$), they are not

[4]The SCR already encodes information about changes in electrodermal response, and so it does not need to be differentiated, since it already encodes changes in perspiration.

Piece ID	dA	dV	dHR	SCR
1	4	5	6	4
2	2	8	6	1
3	1	1	1	2
4	2	4	5	2
5	0	0	1	0
6	13	2	10	4
7	13	11	9	2
8	16	3	14	3
9	2	10	6	2

Table 5.7: Number of peaks in each variable per music.

statistically significant.

variables	dA	dV	dHR	SCR
dA	-	-	-	-
dV	0.116	-	-	-
dHR	0.909*	0.277	-	-
SCR	0.558	-0.019	0.580	-

Table 5.8: Correlations between number of peaks in each variable. * $p < 0.001$

The co-occurrence of peaks (synchronisations between strong changes) in physiological variables and self report was also analysed. In Figures 5.15 to 5.23, the first order differentiation of arousal (dA), valence (dV) and HR (dHR) is shown together with the SCR. The strong changes in each variable are also indicated in the plots. The search for synchronisations focused on the physiological peak changes that precede or coincide with self report events. The physiological events are considered to follow the music stimulus. Because the delay between the music stimulus and the participant response is expected to be in the order of 1 to 5s (Schubert & Dunsmuir, 1999), the time window considered for classifying events as associated was the same. Physiological events synchronised with self report events, or preceding these events up to 5 seconds, are considered to be associated. The results are shown in Table 5.9.

These results show that strong changes in heart rate were followed by self

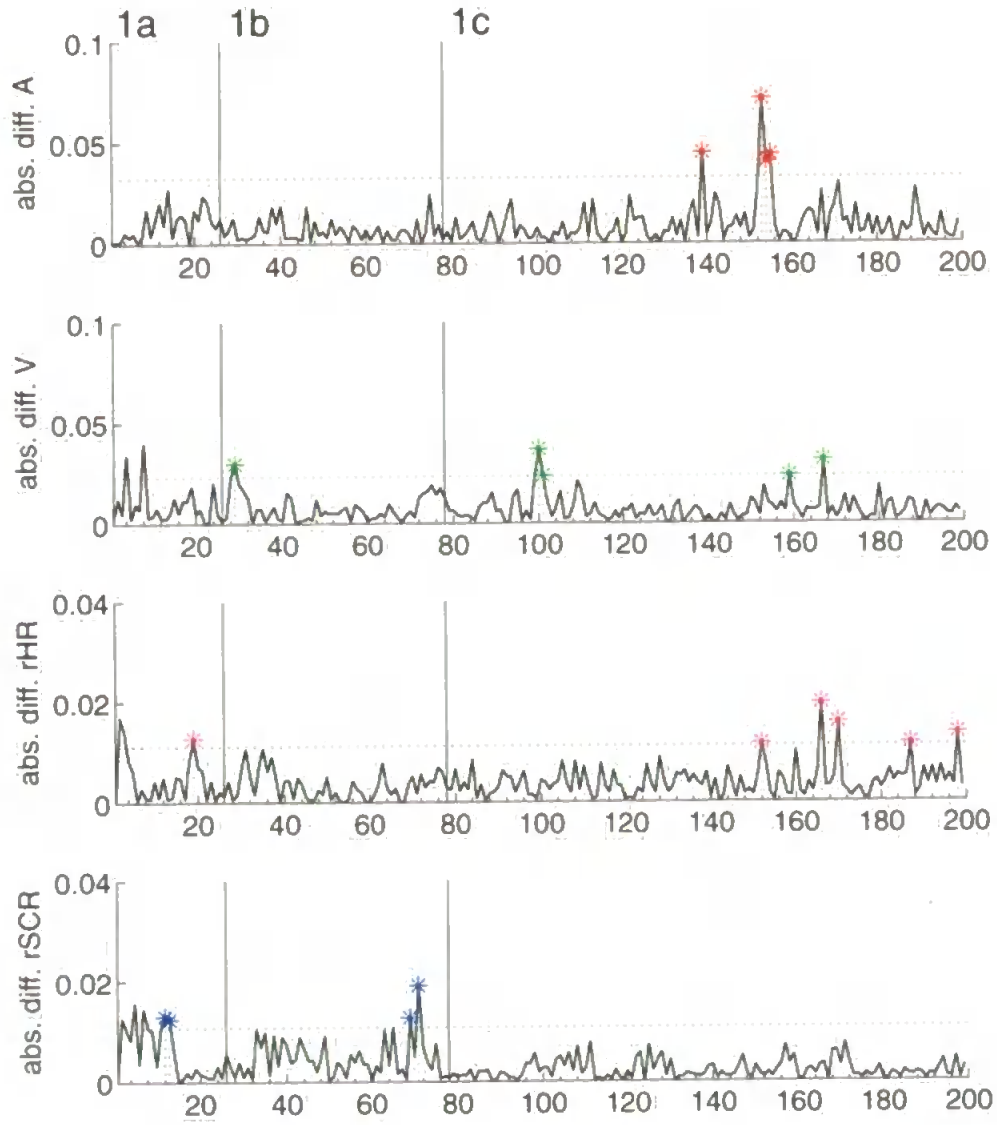


Figure 5.15: Albinoni, *Adagio* (Piece 1): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars.

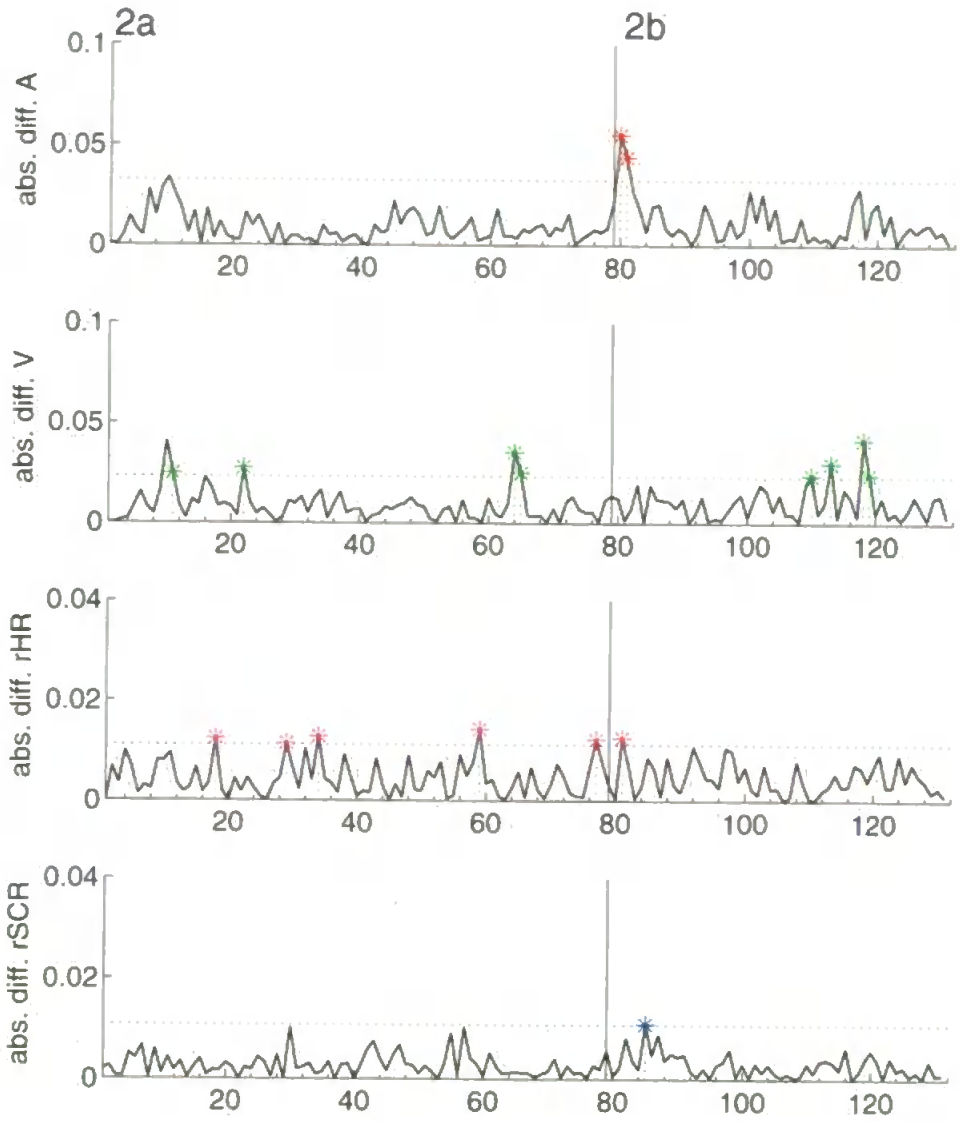


Figure 5.16: Grieg, *Peer Gynt Suite No. 1* (Piece 2): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars.

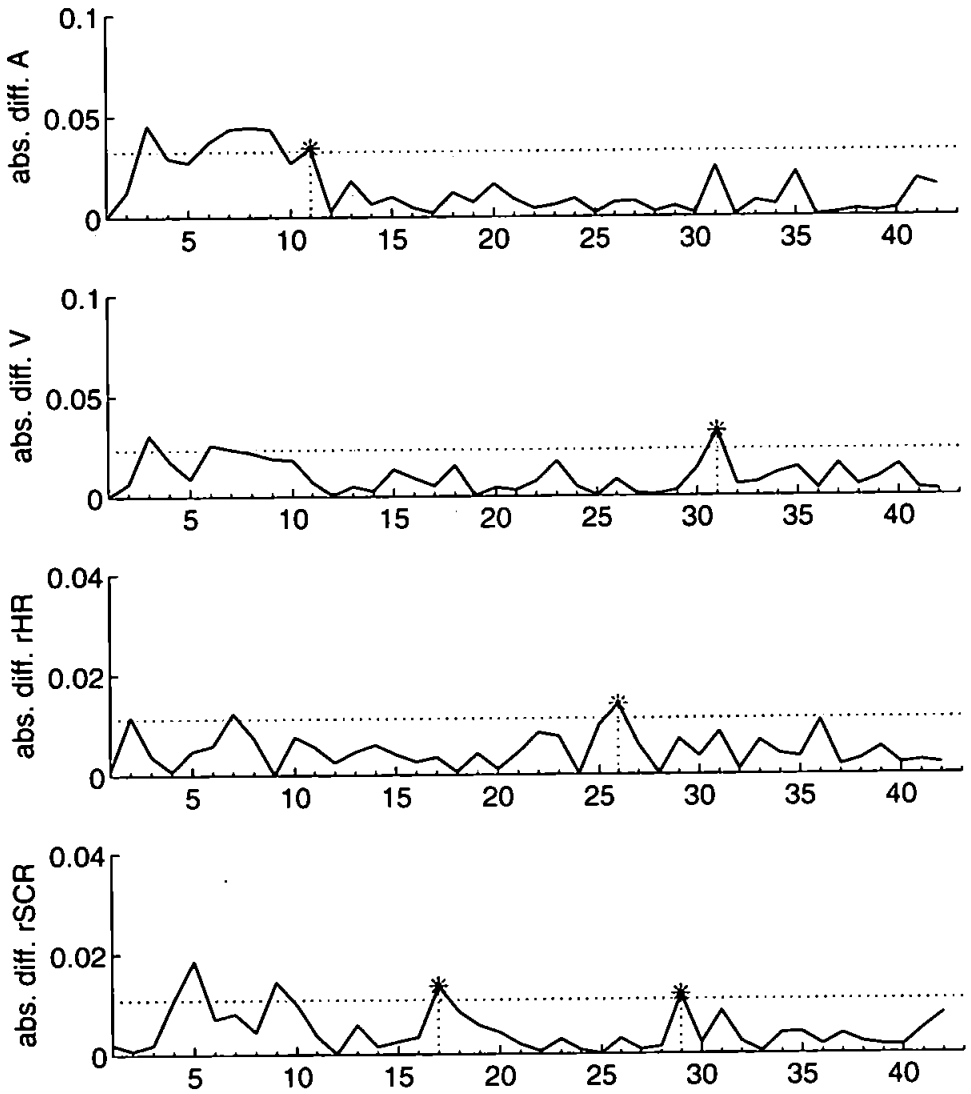


Figure 5.17: Bach, *Prelude and Fugue No. 15* (Piece 3) : first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars.

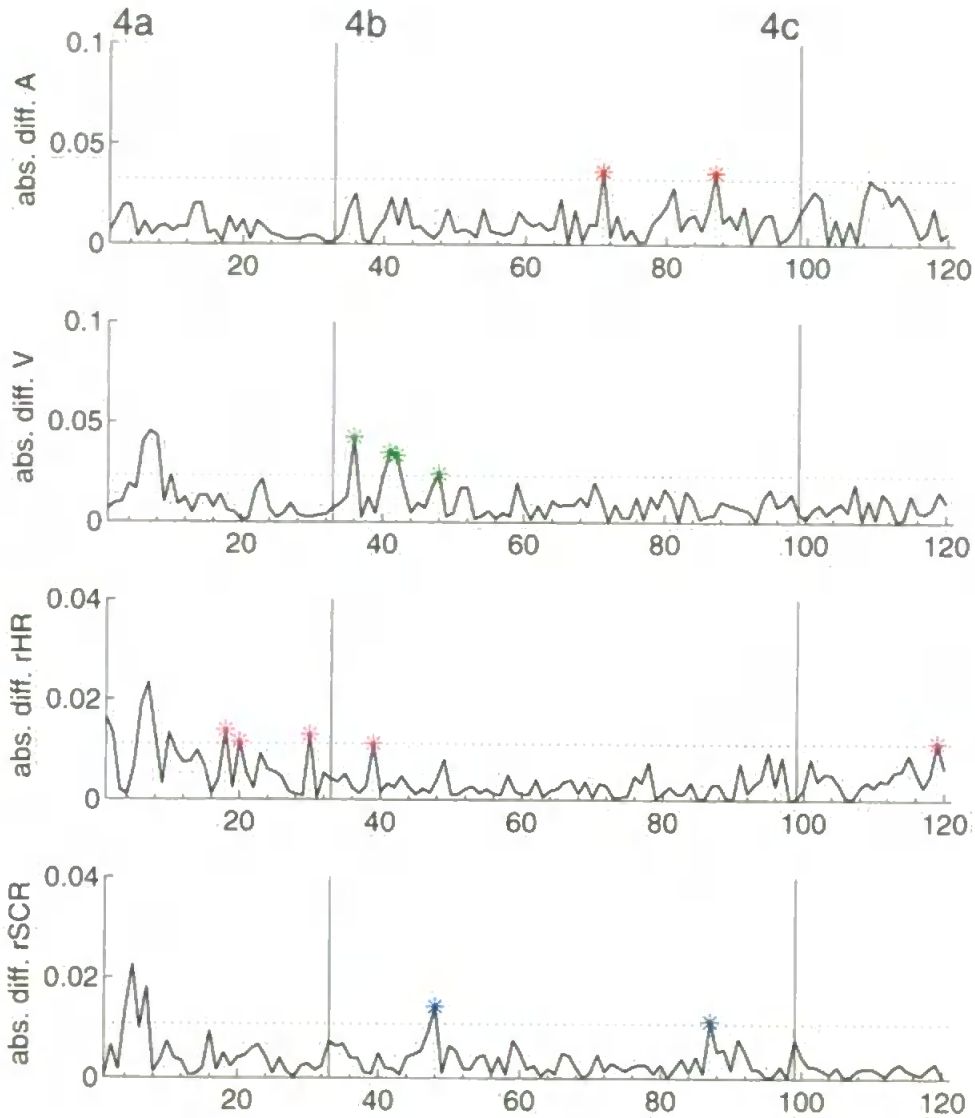


Figure 5.18: Beethoven, *Romance No. 2* (Piece 4): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars.

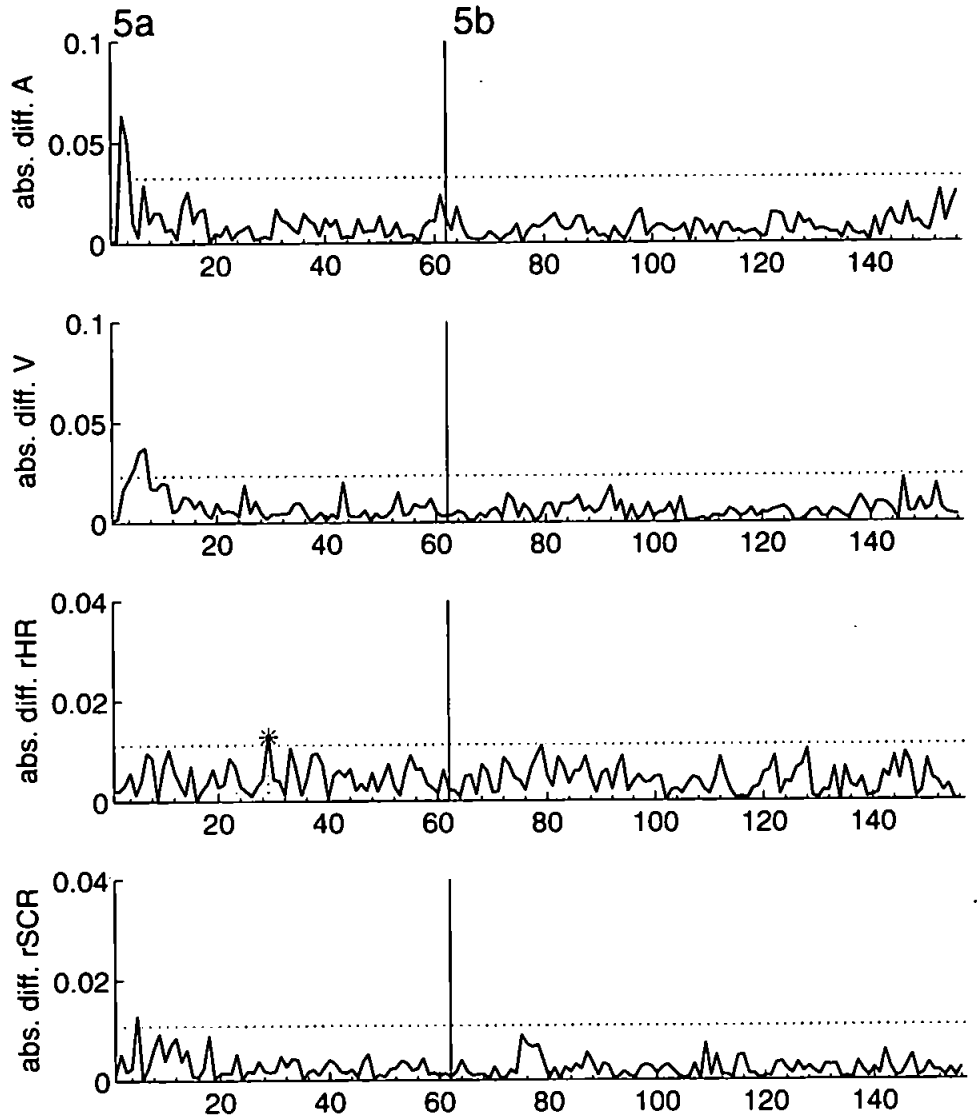


Figure 5.19: Chopin, *Nocturne No. 2* (Piece 5): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars.

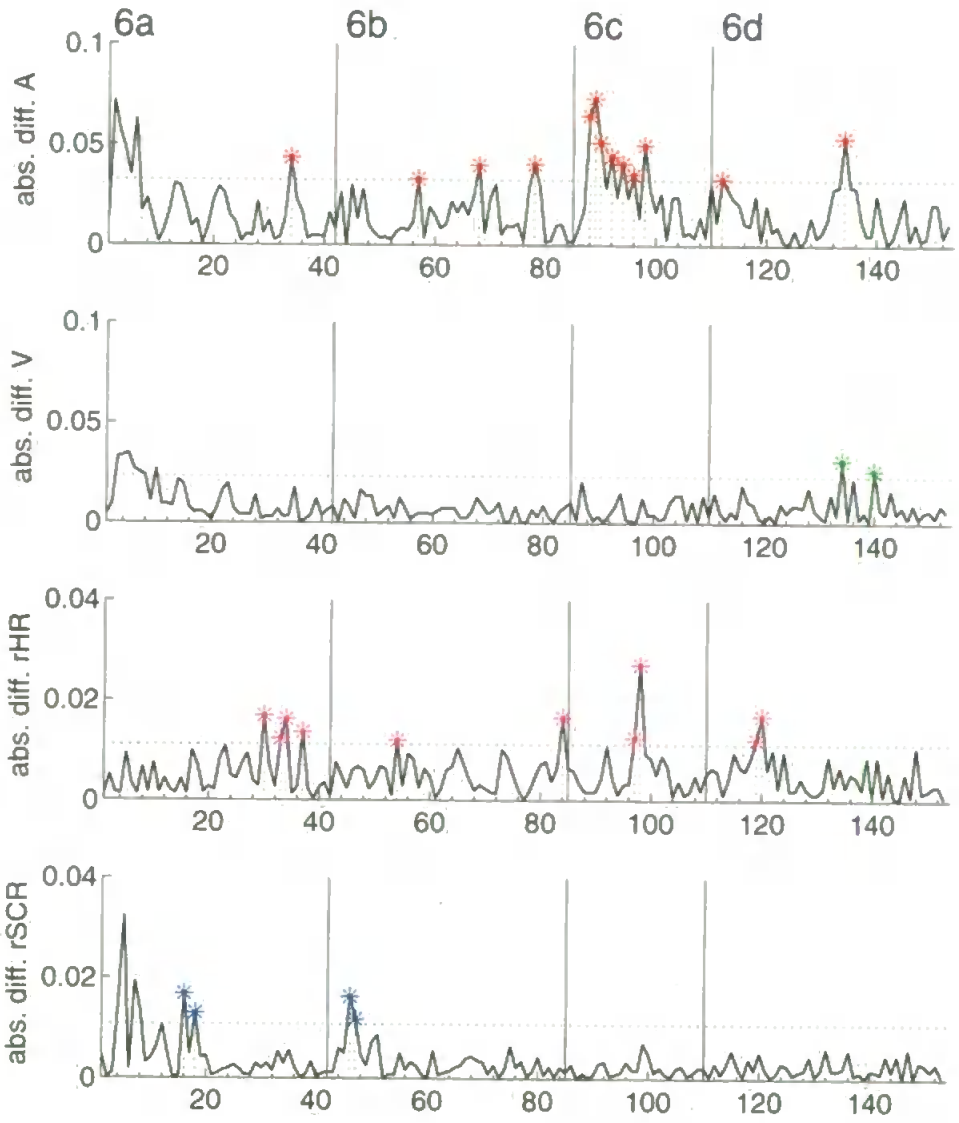


Figure 5.20: Mozart, *Divertimento* (Piece 6): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars.

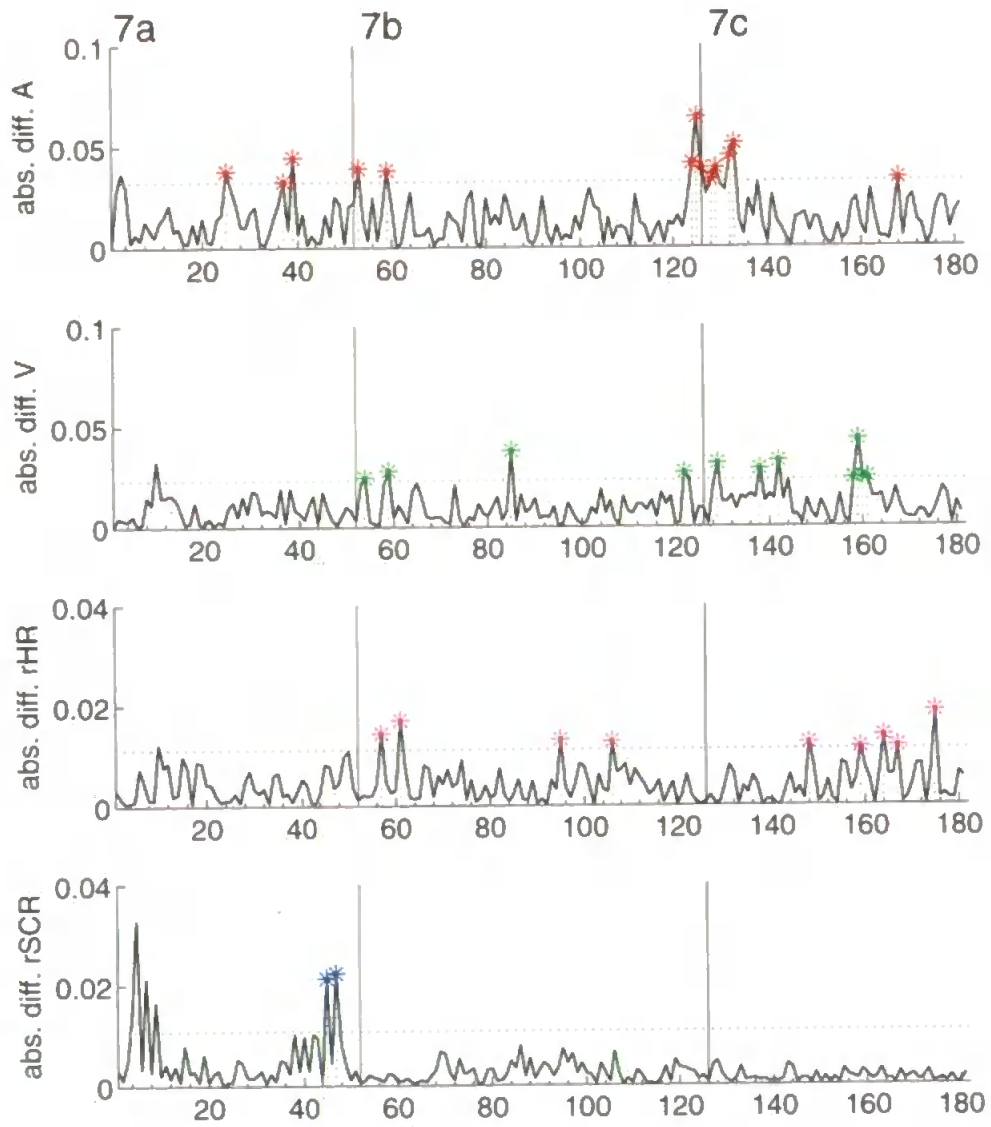


Figure 5.21: Debussy, *La Mer* (Piece 7): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars.

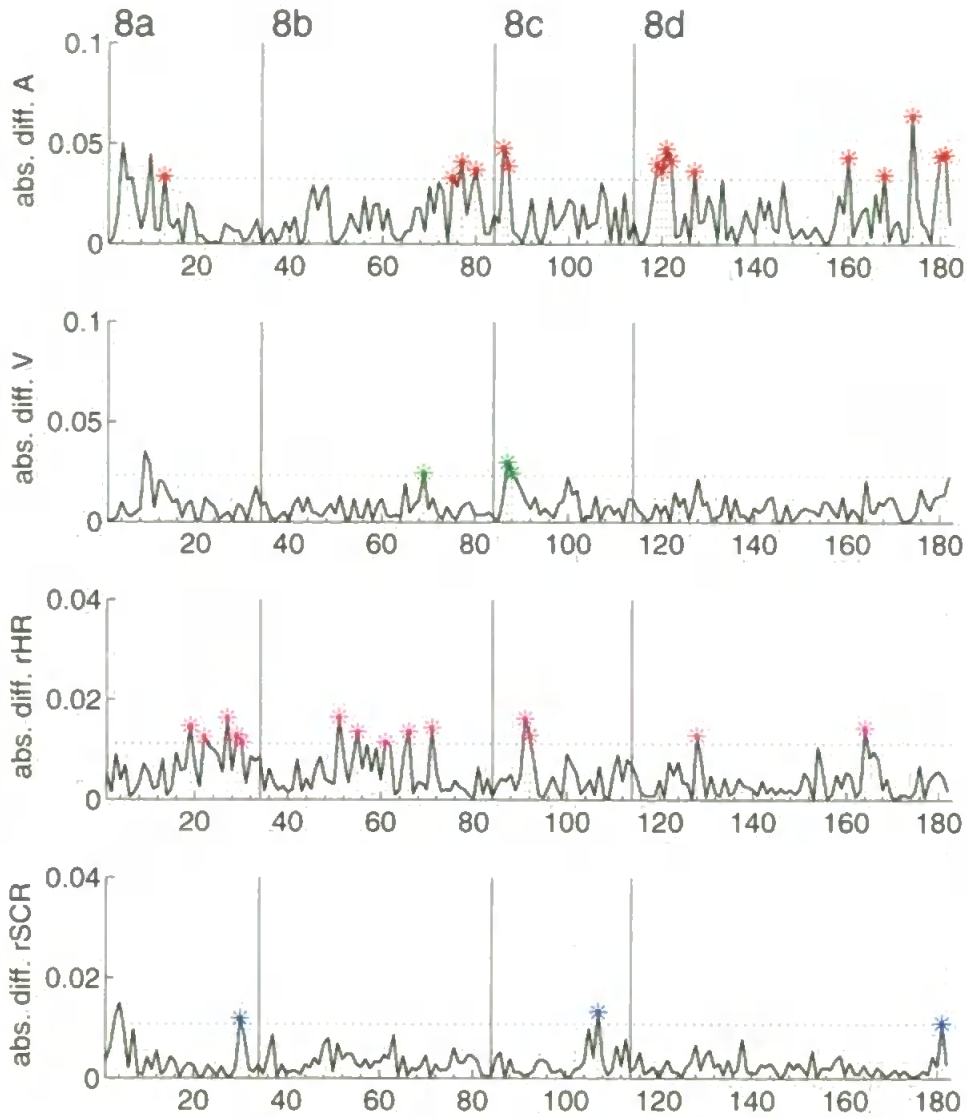


Figure 5.22: Liszt, *Liebestraum No.3* (Piece 8): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars.

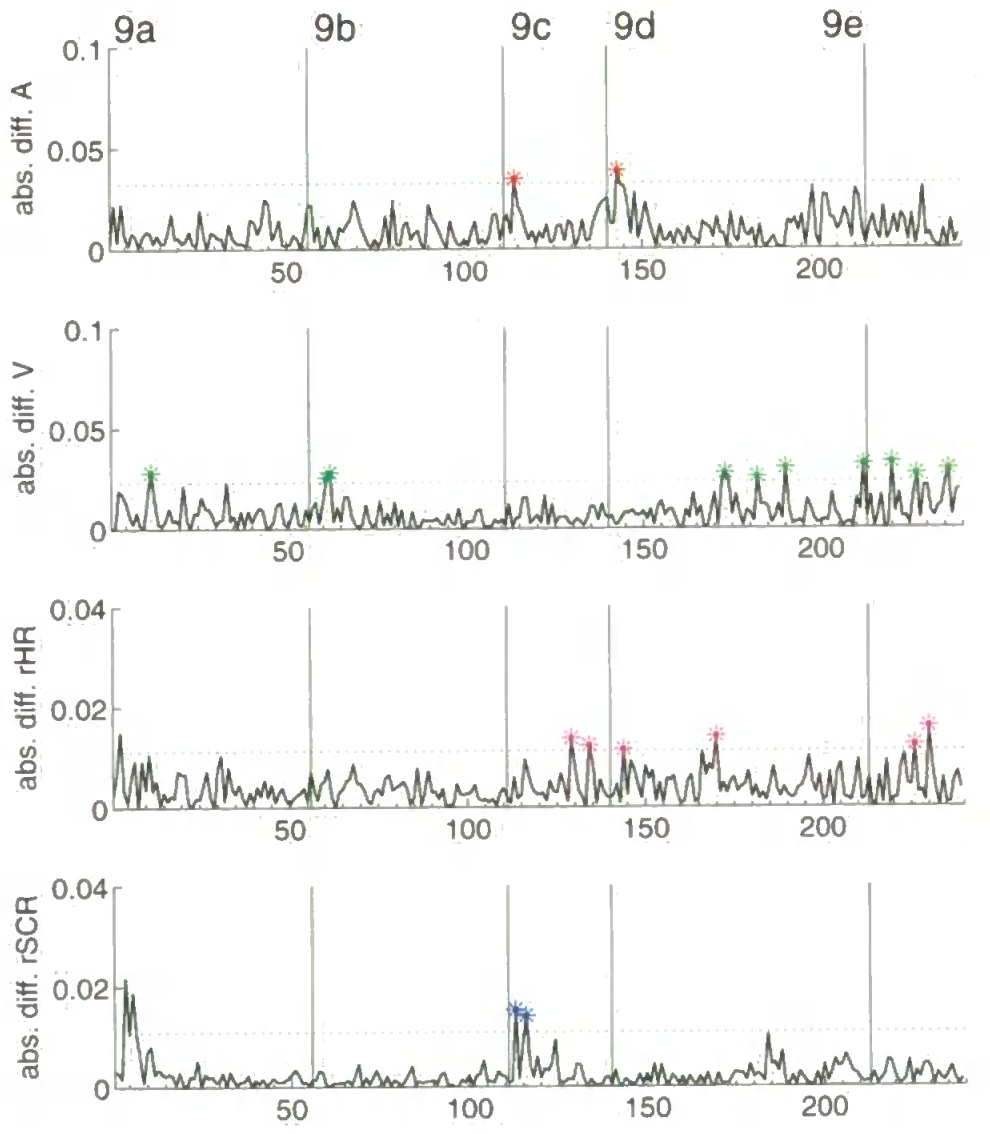


Figure 5.23: Bach, *Partita No. 2* (Piece 9): first order differentiation of Arousal (dA), Valence (dV) and HR (dHR), and SCR. The strong changes in each variable are indicated with coloured stars.

Piece ID	dA-dHR	dV-dHR	dA-SCR	dV-SCR
1	3-1	1-1	0-0	0-0
2	2-2	2-2	0-0	0-0
3	0-0	1-1	0-0	0-0
4	0-0	2-1	1-1	0-0
5	0-0	0-0	0-0	0-0
6	13-6	0-0	0-0	0-0
7	2-3	3-1	1-1	0-0
8	1-1	1-1	1-1	0-0
9	1-1	3-3	1-1	0-0
sum	12-14	13-10	4-4	0-0
perc.	23-24%	30-17%	8-10%	0-0%

Table 5.9: Number of strong changes in self report within 1-5s of physiological events.

report changes 41% of the time: heart rate peaks preceded 23% of the arousal changes and 30% of the total peaks in valence. The skin conductance response only synchronised sporadically with arousal events, suggesting that it may be responsive to musical events with no affective meaning attributed.

5.3.5 Self-report discrimination from sound features and physiological activity

In order to further investigate the existence of specific dimensions of physiological dynamics related to the reported conscious emotional state, a Linear Discriminant Analysis (LDA) was conducted on the segments' mean data.

The LDA is a classic method of classification using categorical target variables (features that somehow relate to or describe the data). Unlike Principle Component Analysis (PCA), in LDA the groups are known or predetermined[5]. The main purpose of using this algorithm is to find the linear combination of features that best separate between classes or object properties. This method

[5]Both methods are very similar because they look for linear combinations of variables which best explain the data; the essential difference consists of the rules for classification (clustering), which are based on distance measures in PCA, while LDA explicitly attempts to model the difference between the classes.

maximises the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability.

The procedure was divided into two different phases. In the first, the data set used for the analysis included both psychoacoustic and physiological variables. The second included only the music ones. The grouping variable was the same for both: the quadrant of the 2DES to which each segment belongs (ranging from 1 to 4). The idea is to assess the level of discrimination in the 2DES by knowing the mean level of each sound features in each segment and the physiological activity. The 2 discriminant functions resultant of each LDA are shown in Fig. 5.24.

The first classification test successfully categorised 81.5% of the test cases. As expected from the multi-variate analysis results, the sound features have great predictive power over self report data. A more interesting result was obtained when the physiological variables were added: the classification success rate increased to 92.6%.

In an attempt to evaluate the direction of interaction between self report and the activation of the autonomic system, another LDA was performed. This time the goal was to find directions of linear interaction on music and self report data, based on the separation in groups differentiated by the level of physiological activation. Groups also varied from 1 to 4 as in the circumplex model of arousal and valence, but HR and SCR, respectively, replaced the axis meanings (e.g. Q_{phys} 1: high HR, high SCR; Q_{phys} 2: low HR, high SCR). The classification in this last test was only successful for 63% of the cases, evidently minor compared with the previous test conditions.

The higher differentiation on the general levels of activation and pleasure by using the physiological dimensions together with psychoacoustic variables, suggests that heart rate and skin conductance response add relevant discriminatory power over the conscious evaluation of the affective response reported by

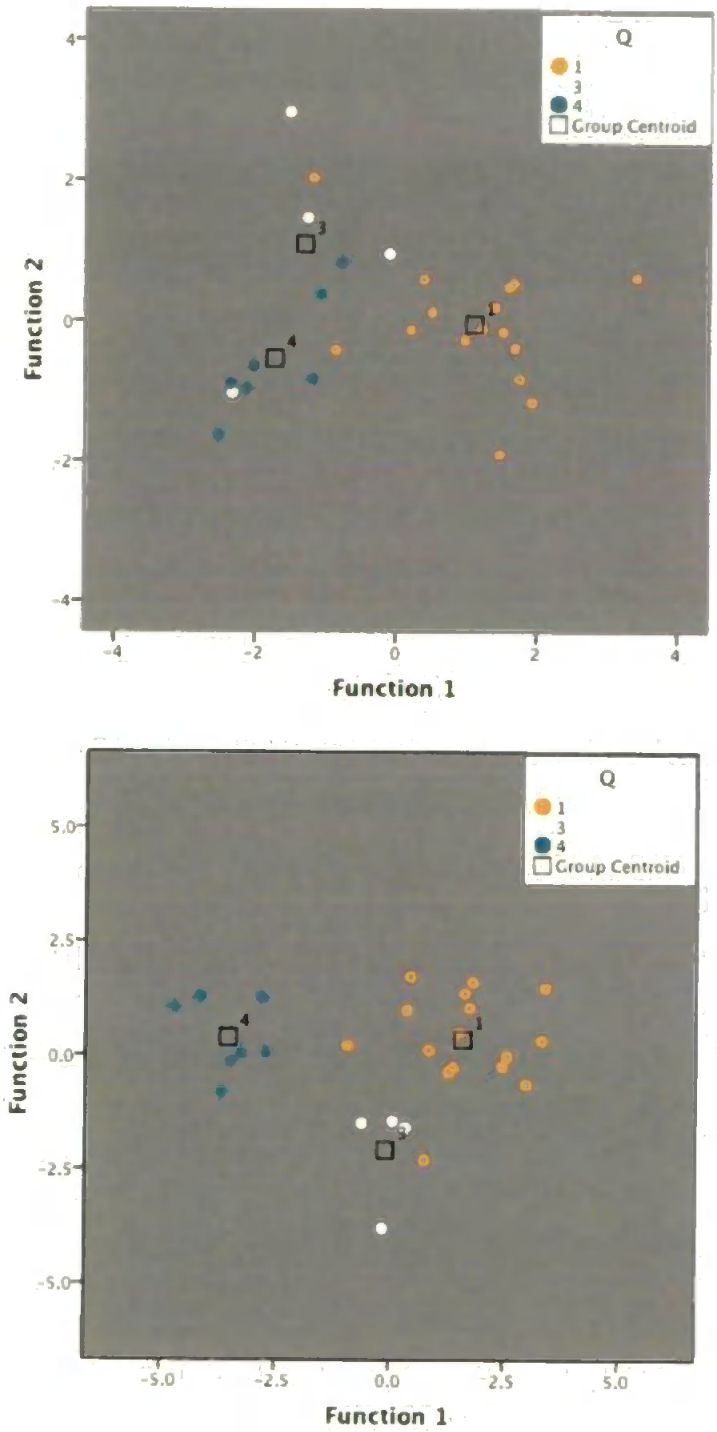


Figure 5.24: Linear Discriminant Analysis: only psychoacoustic variables (left); psychoacoustic and physiological variables (right).

listeners.

5.4 Discussion

For this experiment it was hypothesised that music can alter both the physiological component and the subjective feeling of emotion in response to music, sometimes in highly synchronised ways. The results obtained also support this hypothesis. Throughout this chapter it was shown that loudness, tempo, mean pitch, sharpness and timbral width have a positive effect on both psychological arousal and valence (stronger on the first). The relationships between physiological and psychological data have also shown that arousal and heart rate correlate significantly. The analysis of synchronisations between strong changes in the variables confirmed this relationship and revealed another: almost half of the strong changes in heart rate were followed by changes in reported arousal and valence. Additionally, experimental observations also support this hypothesis. Despite the experimental setup, the focus on music structure and the self report tool, there was evidence that the experiment was successful in evoking genuine emotions in the listeners. Several participants reported the changes in their emotional states while listening to the pieces, one participant cried during the experiment (while listening to pieces 8 and 9) and others have shown evident emotional and behavioural changes (e.g. becoming very talkative, dancing and singing, changes in facial expressions).

This interaction between self report and physiological activity was further investigated with linear discriminant analysis. In 81.5% of the cases, combinations of the sound features derived a successful classification of the general affective value (2DES quadrant) of each segment. By combining sound features and physiological variables, the rate increased to 92.6%. This apparent improvement suggests that physiological cues combined with sound features give a better

description of the self report dynamics. This also suggests self report of emotion is related to physiological activation. Furthermore, 41% of the strong changes in heart rate preceded (up to 4s) strong changes in the subjective feeling component.

Our results are consistent with previous findings reported in the literature. Focusing on the concept of "chills" (subtle nervous tremors caused by intense emotions), Grewe et al. (2007) suggested that strong emotions (or "chills") in response to music are related to structural musical elements. Their analysis supports the claim that important musical factors seem to be harmonic sequences, the entrance of a "voice", and the beginning of new parts (violations of expectancies), as the result of attentive, experienced, and conscious musical enjoyment. In 33 significant general affective events that they analysed, Grewe et al. (2007) identified 7 changes in self report together with an SCR response (21%). Also investigating the nature of "chills" experienced during music listening, Guhn et al. (2007) found that the passages that triggered the highest number of chills also elicited the greatest physiological reactions, namely the largest average increases in heart rate and skin conductance. Their findings also support the idea that physiological changes and chill experiences relate to a number of specific musical characteristics, such as the alternation between a solo instrument and the orchestra, sudden or gradual increase in loudness, or expansions in either low or high register. Rickard (2004) has also suggested that intense responses to music may be differentiated through the skin conductance response, therefore supporting the arousal hypothesis. In another study, Khalfa et al. (2002), also suggests that event related skin conductance responses are sensitive measures of music induced emotions. Results revealed that musical excerpts could induce skin conductance responses that differ according to certain feelings of emotion. For example fear and happiness were associated with higher skin conductance response than sadness and peacefulness, affective groups differentiated by their

arousal level.

Skin conductance responses were found mainly at the beginning of the pieces and responsive to sound events with no relationships with self report changes. As suggested by Rockstroh et al. (1987), it is plausible that certain musical changes may act as orienting reflexes, or reactions to novelty in stimuli, which are characterised by skin conductance response activity (Ben-Shakhar, Gati, Ben-Bassat, & Sniper, 2000). In this experiment several peaks in skin conductance response were found to be elicited in the context of loudness and tempo changes, entrance of instruments, segments division, which can represent changes that act as novelty in the stimulus.

The heart rate was also tested by Krumhansl (1997). The results the study have shown lower heart rate levels related to sad excerpts compared to fear or happy excerpts, contradicting previous studies that suggest that this measure tends to be higher for sad and fear rather than for happy (see Zajonc & McIntosh, 1992). Our analysis has shown a positive correlation between the heart rate level and music tempo. The density of strong changes in arousal was also correlated to heart rate changes. Additionally, it was shown that 41% of strong changes in heart rate preceded 53% of the strongest changes in arousal and valence, suggesting a link between physiological and psychological responses, with an affective meaning.

The data obtained in this experimental study is used in the next chapter to analyse the relationships between the dynamics of sound features, subjective feeling and physiological activation, using a computational model.

Chapter 6

An Extended Neural Network Model of Musical Emotions: The Role of Physiological Arousal

In Chapter 4, it was suggested that spatiotemporal connectionist networks (Kremer, 2001) offer an ideal platform for the investigation of the dynamics of affective responses to music. Following that claim, an Elman neural network was used to model continuous measurements of affective responses to music (arousal and valence), based on a set of psychoacoustic components extracted from the music stimuli (sound features). A significant part of the listeners' affective response was predicted from the psychoacoustic properties of sound, suggesting that these sound features (to which Meyer (1956) referred as "secondary" or "statistical" parameters), encode a most of the information that permits to approximate human affective responses to music. The simulation analysis has also shown consistent and meaningful relationships between the sound features, arousal and valence (see Section 4.3.4).

In order to improve the description of emotion, an experiment was conducted (as described in Chapter 5) to retrieve information about the psychological and

physiological components of emotion, while listening to classical music. One of the main goals of adding physiological cues was to improve the description of the affective experience with music. Participants' heart rate (HR) and skin conductance response (SCR) were the variables measured. The aim of this chapter is to reproduce the model created in Chapter 4, verify its validity and to analyse the relationships between physiological activity with the affective response.

The first part of this chapter presents simulation experiments on two new neural network models of emotional responses to music. "Model 1" only uses sound features as inputs to the model, using the same architecture of the model shown in Chapter 4, and the self report data obtained in the experiment described in Chapter 5. Three new sound features are also tested in comparison with those used in the Chapter 4 model. An important difference between Korhonen's data and the experiment presented in this thesis, is that participants were asked to report the emotion "felt" while listening to the music rather than the "perceived" emotional value of the music (Gabrielsson, 2002). By applying the model developed in Chapter 4 to the new data it is possible to investigate the effect of this factor on the model's performance. "Model 2" evaluates another neural network that includes not only the sound features as inputs, but also the physiological variables. This attempts to predict the dynamics of arousal and valence simultaneously from both sound features and physiological responses, and to investigate the specific contribution of the physiological input (peripheral feedback). In the final part of the chapter, Model 1 and Model 2 are compared and the best network is analysed in detail to infer the transformation rules used to predict affective responses to music.

6.1 Simulation Methodology

The experimental data for this simulation study was obtained in the experiment described in Chapter 5. The self report data includes the arousal and valence emotion descriptors of nine selections of classical music, obtained from 39 participants (see Section 5.2.1). Using a continuous measurement framework, emotion was represented by its valence and arousal dimensions. The physiological measures measured were heart rate and skin conductance level (see Section 5.2.5 for further details).

6.1.1 Music pieces

Table 6.1 shows the music pieces used in the experiment. A full description of the pieces used is shown in Section 5.2.3 of this thesis.

Piece ID	Alias	Duration
1	Adagio	200s
2	Grieg	135s
3	Prelude	43s
4	Romance	123s
5	Nocturne	157s
6	Divertimento	155s
7	La Mer	184s
8	Liebestraum	183s
9	Chaconne	240s

Table 6.1: Aliases for the pieces used in the experiment. See Table 5.1 for further details

6.1.2 Psychoacoustic encoding (model input data)

The sound features that quantify the music stimuli into sound (psychoacoustic) features are shown in Table 6.2.

There are three new sound features in respect to the model presented in Chapter 4. The melodic pitch (mP) was not included in the previous model

Measure	Sound feature	Alias
Dynamic Loudness	Loudness	L
Beats-per-minute (bpm)	Tempo	T
Power Spectrum Centroid	Mean Pitch	P
Multiplicity	Texture	Tx
Mean STFT Flux	Pitch Variation	Pv
Melodic Pitch	Contour	mP
Sharpness	Timbre	S
Timbral Width	Timbre	TW
Roughness	Roughness	R

Table 6.2: Psychoacoustic variables considered for this study.

because there was no algorithm available at the time to extract the melodic line of polyphonic sounds. The measure used in this thesis estimates the contour of the main melodic line of polyphonic sounds, which makes it suitable for the pieces used, since several are orchestral pieces. Timbral Width (TW) was already tested in the first simulation experiments, but was not included in the final model. Sharpness was chosen instead to represent timbre, because it showed a better performance. Nevertheless, as discussed in Section 5.2.4, timbre is a multidimensional quantity, and so it was decided to test it again in the new model in order to improve the description of this quantity. The rationale behind the choice of timbral width (referred to as Ti_2 in Chapter 4), was the fact that, when included in the input set, it led to the second best model performance ($rms_{Ti_2} = 0.089$). Additionally, it has also shown the best valence predictions for novel music ($rms_{Ti_{2-testV}} = 0.080$), when compared with the remaining features (see Section 5.2.4 for further details). Finally, Roughness (R) is also a new sound feature used in the new model. Roughness is a basic psychoacoustic sensation for rapid amplitude variations, which reduces the sensory pleasantness and the quality of noises. Like the mP measure, R was not included in Korhonen's sound features set or in the simulation experiments in Chapter 4. The remaining variables are the same as those described in Section 4.1.2. A visualisation of the individual time series for each piece is included in Appendix F.

6.1.3 Simulation procedure

The simulation methodology for this model is similar to the one presented in Chapter 4. The normalised sound features are the inputs for the model, each corresponding to a single input node of the network. The output layer is again formed by 2 nodes: one representing arousal and the other valence. The dimension of the hidden layer (number of nodes) will be tested again, although previous analysis has shown five to be the ideal number of hidden nodes. In this chapter, the physiological features will also be used as inputs for the model in some of the simulations.

The "training set" (collection of stimuli used to train the model) includes 5 of the pieces used in the experiment (pieces 1, 4, 5, 6 and 8). The "test set" (novel stimuli, unknown to the system during training, that test its generalisation capabilities) includes the remaining 4 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 logical basis behind this decision is the fact that, for the model to be able to predict the emotional responses to novel pieces in an ideal condition, it is necessary that it has been exposed to widest range of values possible. The selection was made based on the observation of Figure 5.13. Figure 6.1 shows the areas covered by the pieces grouped by training and test sets, which correspond to the overlap of all the pieces belonging to each data set (shown in Figure 5.13). Both sets contain extreme values in each variable.

At each training iteration, the task (t) is to predict the next (t+1) values of arousal and valence. The "teaching input" (or target values) are the average A/V pairs obtained experimentally. The range of values for each variable (sound features, self report and physiological variables) was normalised to the range between 0 and 1 in order to be used with the model. The learning process didn't change from the model described in Chapter 4 and it was implemented using a standard backpropagation technique (Rumelhart et al., 1986).

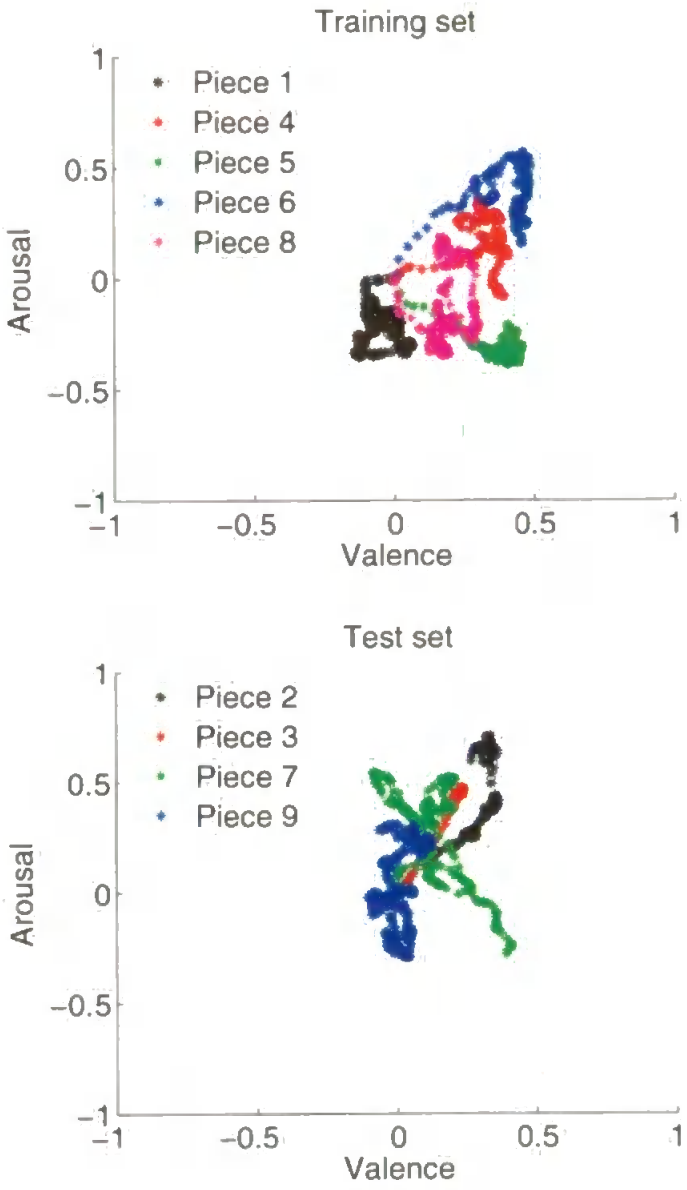


Figure 6.1: Areas covered by the pieces grouped by training (a) and test (b) sets. In both plots, the colours of each cloud of points correspond to all the Arousal/Valence pairs from each piece (see legend). Each point corresponds to the specific location on the 2DES of the Arousal and Valence values on a second by second basis.

For each replication of the simulations[1], the network weights were initialised with different values randomly 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). Each trial consisted of 80000 iterations of the learning algorithm. The training stop point was estimated *a posteriori* by calculating the number of training iterations which minimise the model outputs error for both training and test sets. This is a fundamental step to avoid the overfitting of the training set. During training the same learning rate and momentum were used for each of the 3 connection matrices. The learning rate was set to 0.075 and the momentum to 0.0 for all trials. The rms (root mean square) error will be used to quantify the deviation of the model outputs from the values observed experimentally. The model performance will be assessed with the linear (r) correlation coefficient.

The modelling process will be divided in two main groups of simulations. The first part of the study reproduces the model presented in Chapter 4, which is evaluated in its architecture (sound features and hidden layer size) with the new experimental data. This part (Model 1) is strictly focused on the sound features and self reported arousal and valence variables. The second part of the study (Model 2), evaluates the role of the inclusion of physiological cues as inputs to the model (the peripheral feedback hypothesis).

6.2 Model 1: Modelling musical emotions with sound features

Model 1 is a reproduction of the model described in Chapter 4 with new experimental data and new music pieces. It only considers sound features as inputs for the model and the self report of emotion (arousal and valence) as

[1]Each model will involve a set of simulations. Each simulation consists of a set of replications (fifteen in this chapter) in which the same simulation is repeated with different initial conditions (randomised weights).

the outputs. The main difference between the two models relies on the pieces used and the type of self report obtained. In this model the experimental data (presented in Chapter 5) contains the self report of emotions "felt". The aim is to predict the dynamics of felt emotions from the dynamics of psychoacoustic patterns of each piece.

6.2.1 Hidden layer size

Like in the previous model, the first step was to find the ideal size for the hidden layer. Seven simulations (fifteen trials each) were run using the model shown in Figure 6.2. In each simulation the number of hidden nodes was set to a different value (varying from 1 to 8). The rms errors for each output and their average across training and test data sets are shown in Table 6.3.

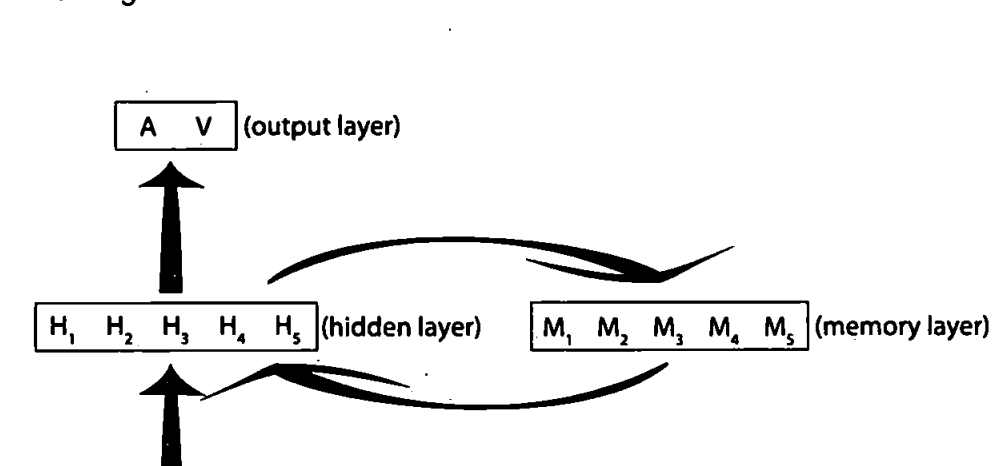


Figure 6.2: Neural network architecture and units identification for Model 1: Input units - sound features (T, Tx, L, P, S and Pv); Hidden units - H_1 to H_5; Memory (context) units - M_1 to M_5; Output units - Arousal (A) and Valence (V).

The smallest deviations from the experimental data correspond to the model with 5 hidden nodes, which shows the best performance for both training and

Sim. ID	Hid. Layer size	Train rms A-V	Test rms A-V	Average rms Train	Test
M_1H_2	2	0.093-0.070	0.104-0.089	0.082	0.096
M_1H_3	3	0.085-0.067	0.091-0.076	0.076	0.084
M_1H_4	4	0.083-0.064	0.088-0.075	0.073	0.082
M_1H_5	5	0.073-0.062	0.083-0.076	**0.067**	**0.080**
M_1H_6	6	0.080-0.066	0.085-0.076	0.073	**0.080**
M_1H_7	7	0.079-0.069	0.088-0.077	0.074	0.083
M_1H_8	8	0.076-0.064	0.097-0.085	0.070	0.091

Table 6.3: Model 1: average rms errors over 15 trials for all 7 simulations with different number of hidden nodes. The mean values across the training and the test data sets are also shown for each model.

test data sets ($rms_{Training-M_1H_5} = 0.067$, $rms_{Test-M_1H_5} = 0.080$). For the following simulations, the hidden (and memory) layer(s) size of Model 1 was fixed to 5 nodes.

6.2.2 Testing new sound features (model inputs)

After establishing the optional number of nodes in the hidden layer size for the model, the model was tested with the inclusion of the additional new inputs (sound features). Three new simulations (15 trials each) were performed varying in each only the inputs to the model. In each simulation a different variable (melodic pitch, roughness or timbral width) was added to the initial set (tempo, texture, loudness, power spectrum centroid, sharpness and mean STFT flux). The average rms errors across all trials of each simulation are shown in Table 6.4.

Sim. ID	Inputs	Training set A	Training set V	Test set A	Test set V	Average
M_1I_0	L, T, P, Pv, Tx, S	**0.073**	**0.062**	**0.083**	**0.076**	**0.074**
M_1I_1	L, T, P, Pv, Tx, S + mP	0.081	0.064	0.089	0.079	0.078
M_1I_2	L, T, P, Pv, Tx, S + R	0.078	0.069	0.088	**0.076**	0.078
M_1I_3	L, T, P, Pv, Tx, S + TW	0.078	0.069	0.091	**0.073**	0.078

Table 6.4: Model 1: simulations with new sound features; the rms errors correspond to the average of the fifteen trials.

The best performance for all sets and both outputs was obtained in simulation $M_1 I_0$ (av. $rms = 0.074$). The only case in which this model performed worst than the remaining ones was for the valence predictions of test data set (the best is model $M_1 I_3$). This result is similar to the one obtained for Model 0 in Chapter 4, one of the reasons that led me to retest timbral width (it has shown also a good performance for the valence predictions to new music). Nevertheless, the global performance is comparatively worst, and so none of the test variables was included in the final model. The final architecture for Model 1 includes same 6 inputs as in Model 0 (tempo, texture, loudness, power spectrum centroid, sharpness and mean STFT flux), 5 hidden and 5 memory nodes, and 2 outputs (arousal and valence).

6.2.3 Model performance

The best network (lowest rms errors for both output variables) of Model 1 was chosen from the initial pool of tests. Table 6.5 shows the rms error and the linear correlation coefficient (r), used to describe the deviation and similarity between the model outputs and the experimental data.

Piece	rms		r		Set
	A	V	A	V	
1	0.058	0.033	0.080	0.252	Train
2	0.051	0.028	0.904*	0.897*	Test
3	0.039	0.014	0.897*	0.966*	Test
4	0.058	0.035	0.823*	0.839*	Train
5	0.069	0.084	0.807*	0.933*	Train
6	0.039	0.066	0.769*	0.897*	Train
7	0.095	0.058	0.775*	0.355*	Test
8	0.095	0.026	0.880*	0.599*	Train
9	0.043	0.094	0.960*	0.329*	Test
av.	0.061	0.049	0.766	0.618	

Table 6.5: MODEL 1: rms errors and r coefficient, per variable, for each music piece for best trial of simulation $M_1 I_0$. * $p < 0.0001$

The model responded with low rms error for both variables and almost all pieces. The exceptions are, for arousal predictions, pieces 7 and 8, and for valence, piece 9. The average rms across all pieces was nevertheless low. The linear correlations between experimental data and model predictions were considerably high ($r(Arousal) = 0.766$ and $r(Valence) = 0.618$). Nevertheless, the pieces with the highest rms for the arousal output (pieces 7 and 8) show a high linear correlation coefficient with the experimental data ($r_7(Arousal) = 0.775$ and $r_8(Arousal) = 0.880$). Instead, Piece 9 (the one with the highest rms error for valence) shows a lower correlation coefficient (although it is still statistically significant). Additionally, no significant linear correlations were found for piece 1.

6.2.4 Comparison with the previous model

The model presented in Chapter 4 (Model 0) is identical to the one presented in this chapter, which permits to compare their performance. Table 6.6 shows the values of the average rms errors and r correlation coefficient averaged over all pieces.

Model	Av. rms		Av. r	
	A	V	A	V
Model 0 (Chapter 4)	0.056	0.059	0.743	0.542
Model 1 (Chapter 6)	0.061	0.049	0.766	0.618

Table 6.6: Comparison between average rms errors and r correlation coefficient for Model 0 (Chapter 4) and Model 1 (this chapter).

The error for arousal was higher for Model 1 ($rms_A = 0.061$) than for Model 0 ($rms_A = 0.056$). On the contrary the model performance for valence was substantially improved from Model 0 ($rms_V = 0.063$) to Model 1 ($rms_V = 0.049$). Although the arousal error was higher, the linear correlations between experimental data and the model predictions were improved especially for valence.

Both models have similar performances, nevertheless Model 1 performed better. Although this difference is small, there are at least two possible readings of this result: either the music pieces used in Model 1 are "easier" to model, or the fact that the data for that model was obtained by asking participants the emotion they feel (rather than the one thought to be expressed by the music). In both cases further experiments with human participants would be required to test these two hypotheses.

6.3 Model 2: modelling musical emotions with sound features and physiological cues

This simulation experiment aims to integrate physiological cues into the model presented in the previous section (Model 1). As discussed in Chapter 2, the "component process model" views emotion as a construct of coordinated changes in physiological arousal, motor expression and subjective feeling, which may be highly synchronised to adapt in an optimal way to the eliciting circumstances. Such a view has received support from consistent evidence about the relation between affective states and bodily feelings making use of physiological measurements (e.g. Harrer & Harrer, 1977; Khalfa et al., 2002; Krumhansl, 1997; Rickard, 2004). Although evidence of an emotion specific physiology was never found (Ekman & Davidson, 1994; Cacioppo et al., 1993), research on peripheral feedback provides evidence that body states can influence the emotional experience with music (Dibben, 2004; Philippot et al., 2002). Peripheral feedback has also been considered to be able to change the strength of an emotion even after this has been generated in the brain (Damasio, 1994). The model presented in this section (Model 2) evaluates a new architecture that includes physiological cues. This is an attempt to predict the dynamics of arousal and valence from both sound features and physiological

responses, and to investigate the specific contribution of the physiological input (peripheral feedback).

6.3.1 Hidden layer size

Given the new input data, a new set of simulations was carried out to establish the ideal size of the hidden layer. Figure 6.3 shows the new model architecture. In a set of seven simulations (fifteen trials per simulation) the number of hidden nodes was varied. Each trial consisted of 80000 iterations of the learning algorithm. The input set includes heart rate (HR) and skin conductance response (SCR) plus the 6 key sound features included in Model 1 (tempo, texture, loudness, power spectrum centroid, sharpness and mean STFT flux). The hidden layer size varied from 2 to 8 hidden units in each simulation. The rms errors for each output and their averages across the training and testing sets are shown in Table 6.7.

Sim. ID	Num. Hid. Units	Train rms A/V	Test rms A/V	Average rms Train	Test
M_2H_2	2	0.093/0.072	0.093/0.081	0.083	0.087
M_2H_3	3	0.082/0.065	0.094/0.077	0.073	0.086
M_2H_4	4	0.080/0.066	0.092/0.076	0.073	0.084
M_2H_5	5	0.070/0.060	0.087/0.075	0.065	0.081
M_2H_6	6	0.072/0.068	0.089/0.072	0.070	0.080
M_2H_7	7	0.070/0.066	0.092/0.073	0.068	0.083
M_2H_8	8	0.074/0.070	0.087/0.072	0.072	0.080

Table 6.7: Model 2: rms errors for 7 simulations with different number of hidden nodes. The mean values across the training and the test data sets are also shown for each model.

From this table, note that the model with 5 hidden nodes shows still the best model performance, even when the physiological variables were added to the inputs.

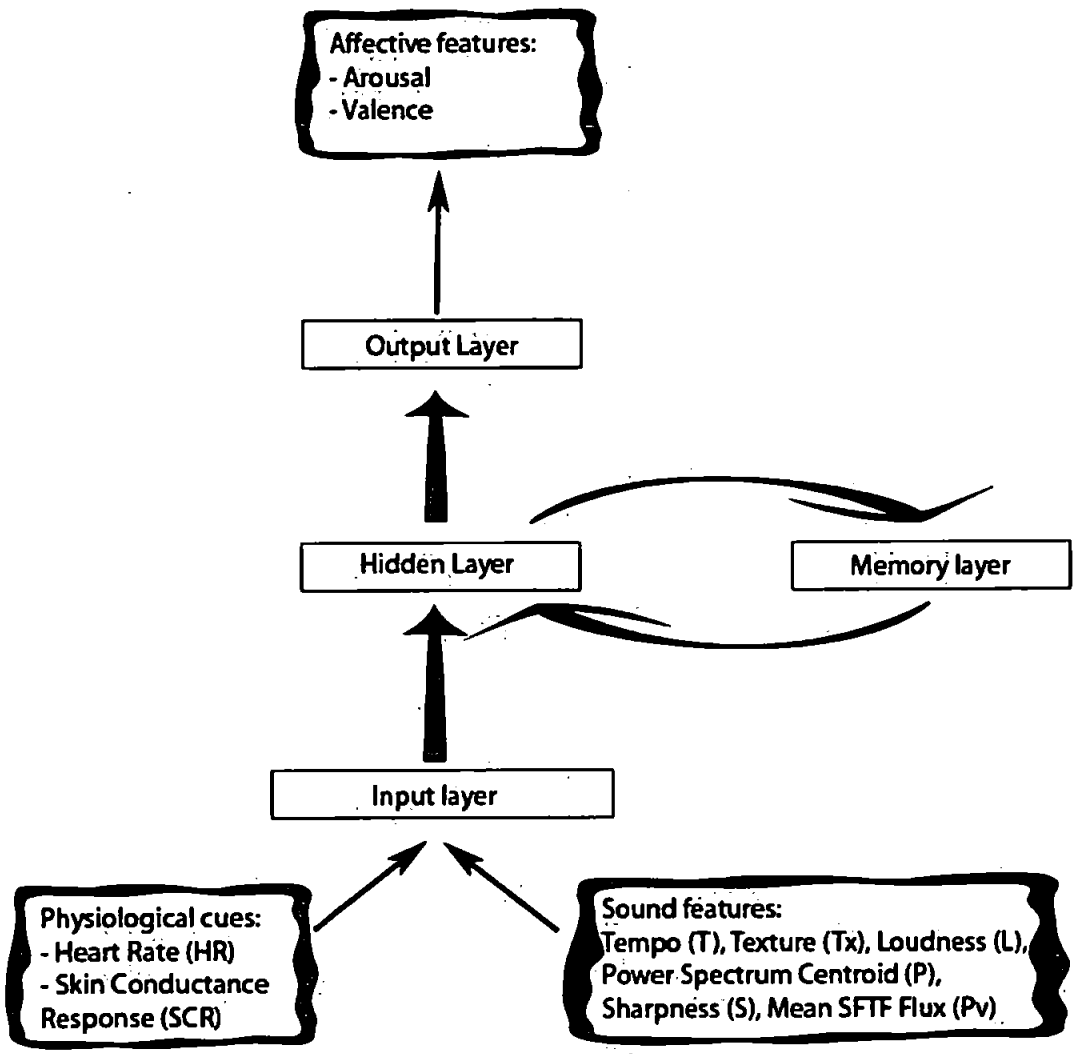


Figure 6.3: Neural network architecture and units identification for Model 2. Inputs: sound features (L, T, P, Pv, Tx, S), physiological variables (HR, SCR). Outputs: Arousal (A) and Valence (V).

6.3.2 New input dimension: physiological variables

In the following simulations 3 different models were tested as reported in Table 6.8. In each simulation, HR and SCR are tested separately and together as inputs to the model.

Sim. ID	Inputs	Training set		Test set		Average
		A	V	A	V	
M_2I_0	L, T, P, Pv, Tx, S + HR	0.077	**0.057**	**0.082**	0.078	**0.073**
M_2I_1	L, T, P, Pv, Tx, S + SCR	0.082	0.067	0.089	0.076	0.079
M_2I_2	L, T, P, Pv, Tx, S + HR, SCR	**0.070**	0.060	0.087	**0.075**	**0.073**

Table 6.8: Model 2: simulations with different combinations of physiological variable inputs (average rms errors over all trials).

The best performance was achieved in simulations M_2I_0 (sound features + extra HR input) and M_2I_2 (sound features + extra HR and SCR inputs). Although the performances of these two models are very similar, the first simulation had a lower error for the test set ($rms(HR) = 0.080$ and $rms(HR, SCR) = 0.081$ - these values correspond to the mean rms errors for both outputs of the test set). Because the addition of SCR does not have any positive impact on the model performance, the set of inputs in simulation M_2I_0 was selected for the final configuration of Model 2.

6.3.3 Model 2 performance

Table 6.9 shows the rms error and the linear correlation coefficient (r) to describe the deviation and similarity between the model outputs and experimental for the best trial of simulation M_2I_0. The values are shown for each piece.

The outputs rms error is low for both variables. The exceptions are, for the arousal output, pieces 7 and 8 (like Model 1). All deviation in the valence output were low (an improvement on Model 1 - see Table 6.5). The linear correlations between experimental data and model predictions were high ($r(Arousal) = 0.864$

	rms		r	
Piece	A	V	A	V
1	0.089	0.034	0.065	0.550*
2	0.036	0.026	0.939*	0.834*
3	0.072	0.051	0.869*	0.949*
4	0.047	0.026	0.884*	0.873*
5	0.063	0.079	0.605*	0.944*
6	0.042	0.060	0.805*	0.892*
7	0.102	0.074	0.815*	0.011
8	0.093	0.027	0.770*	0.627*
9	0.079	0.070	0.813*	0.166
av. Model 2	0.069	0.050	0.864	0.783

Table 6.9: Model 2: rms errors and r coefficient, per variable, for each music piece for best trial of simulation $M_2 I_0$. * $p < 0.0001$

and $r(Valence) = 0.783$) and also correspond to an improvement over Model 1 for both variables (see Table 6.5). No correlations were found in arousal for piece 1, and in valence in for pieces 7 and 9.

In the following section the model presented in this table, is compared to the model shown in the previous section (with only sound features input), in order to evaluate the contribution of physiological cues to the predictions of psychological report of musical emotions.

6.4 Models 1 and 2 comparison

In this section the performance of Models 1 and 2 are compared in terms of their deviation from and similarity to the experimental data. Table 6.10 shows the means values across all pieces for each performance measure between the two models under comparison (Model 1, see Figure 6.2, and Model 2, see Figure 6.3). Figures 6.4 to 6.6 show the Model 1 and Model 2 outputs together with experimental data (target outputs).

In Chapter 5, it was shown that the heart rate level correlated to arousal and

Model	Inputs		rms		r	
	Sound feat.	Phys. feat.	A	V	A	V
1	L, T, P, Pv, Tx, S		0.061	0.049	0.766	0.618
2	L, T, P Pv, Tx, S	HR	0.069	0.050	0.864	0.783

Table 6.10: Comparison between Models 1 and 2: rms errors and linear correlation coefficient (r). These values are the mean values across all pieces for each of the models. The details for each are shown in Tables 6.5 (Model 1) and 6.9 (Model 2).

valence. In the model simulations, the inclusion of HR (Model 2) as an input to the model has had a positive influence on the model performance. For both arousal and valence both the linear correlation coefficient increased significantly. In the following section Model 2 is further analysed.

6.5 Heart Rate and subjective feelings

In the previous sections it was shown that some physiological cues contain relevant information about the dynamics of the affective response to music. Model 2 was able to track the general fluctuations in arousal and valence for the training data but also to predict human responses to another set of novel music pieces. Those results, based on the additional heart rate input, were shown to improve the model that only included the sound features as inputs. In Chapter 4, I introduced a set of analytical methods to reveal some of the strategies the model uses to predict affective responses to music. These methods were used again in the analysis of Model 2: lesioning and correlation analysis. The aim was to infer the dynamics of information flow within the model. In the next section, the discussion will focus on the contribution of heart rate input to the arousal and valence changes.

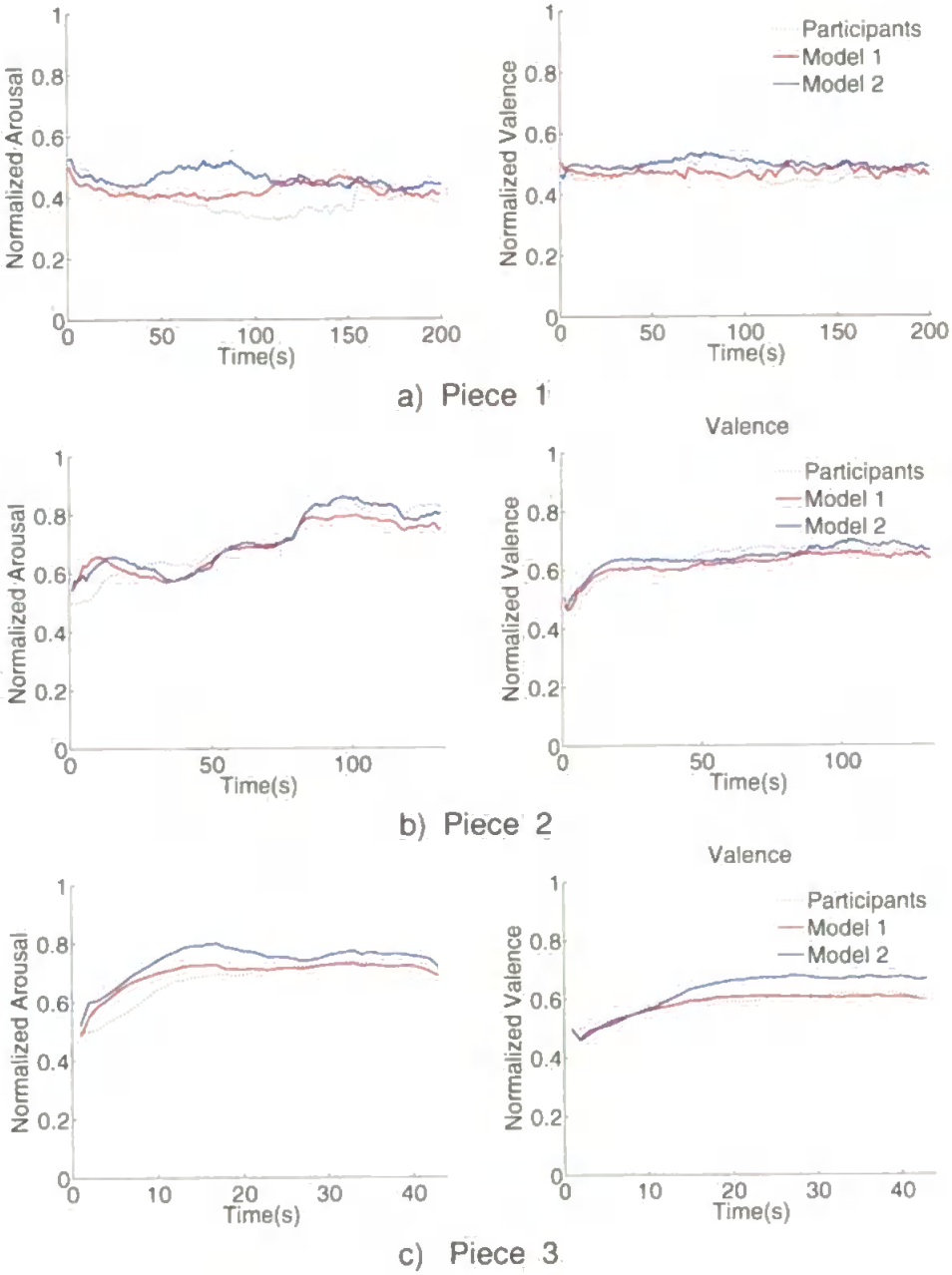


Figure 6.4: Arousal and Valence model outputs compared with experimental data for the training data set: Piece 1 (Albinoni, *Adagio*), Piece 2 (Grieg, *Peer Gynt Suite No. 1*) and Piece 3 (Bach, *Prelude and Fugue No. 15*).

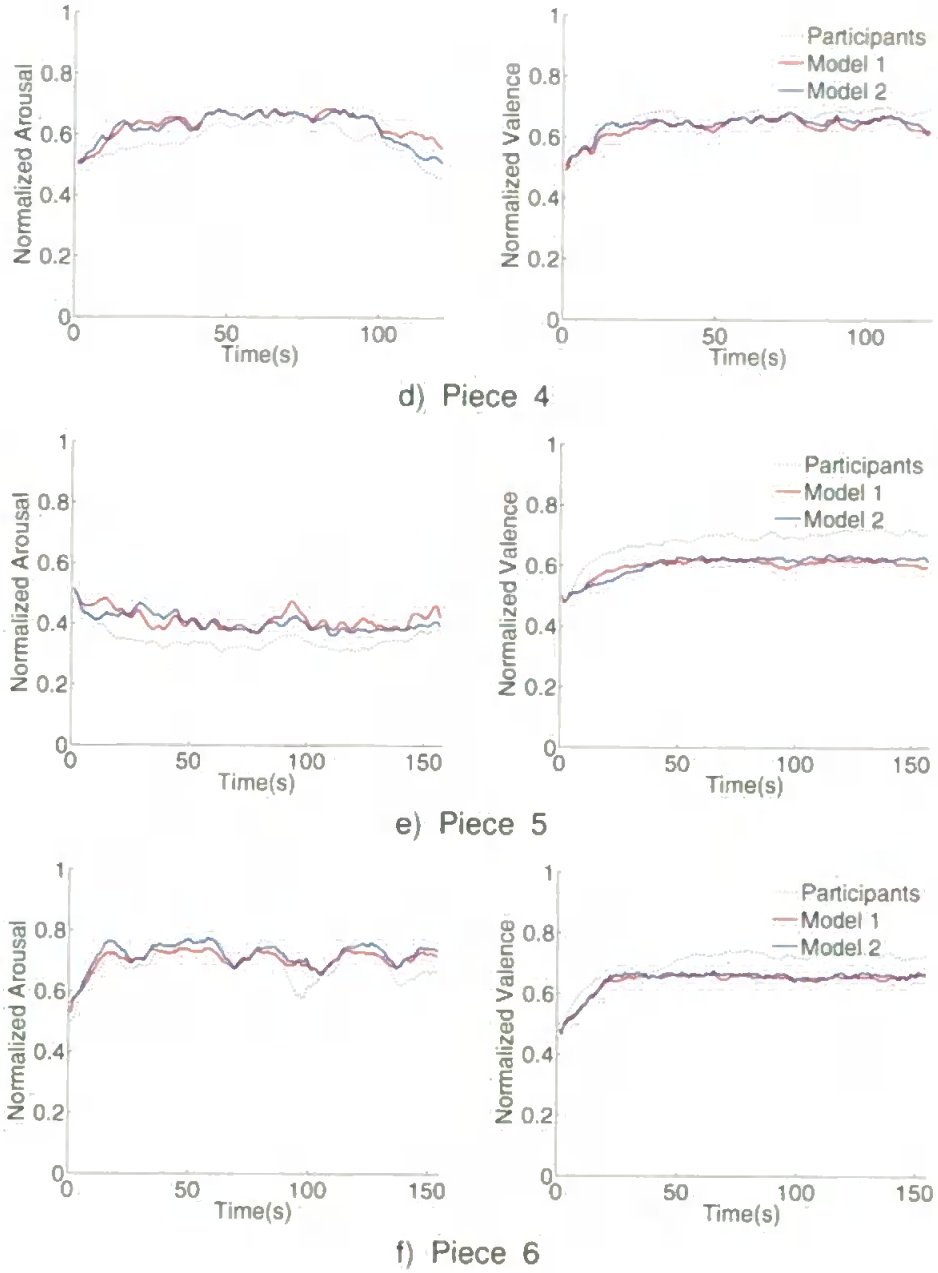


Figure 6.5: Arousal and Valence model outputs compared with experimental data for the training data set: Piece 4 (Beethoven, *Romance No. 2*), Piece 5 (Chopin, *Nocturne No. 2*) and Piece 6 (Mozart, *Divertimento*).

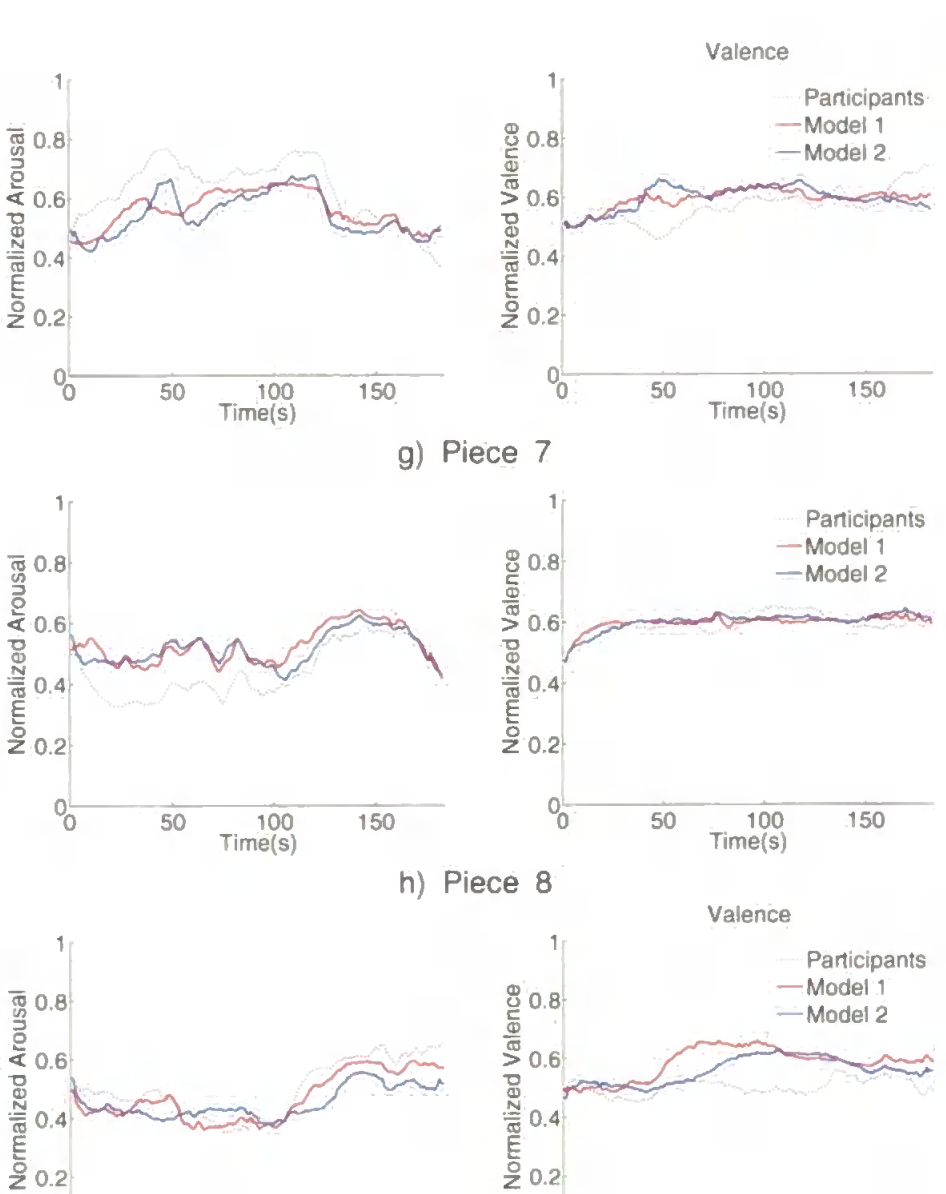


Figure 6.6: Arousal and Valence model outputs compared with experimental data for the training data set: Piece 7 (Debussy, *La Mer*), Piece 8 (Liszt, *Liebestraum No.3*) and Piece 9 (Bach, *Partita No. 2*).

6.5.1 Lesioning tests: long-term memory analysis

Figure 6.7 shows the detailed view of the model architecture. Additionally, the weight matrices are represented in Figure 6.8 where the size of the rectangles is proportional to the weight value (bigger rectangle, bigger weight), and the colour represents the signal of the weight - red for negative, green for positive. All three learned weight matrices in the model are shown (note that the weights from hidden to memory layer are kept constant in ENNs).

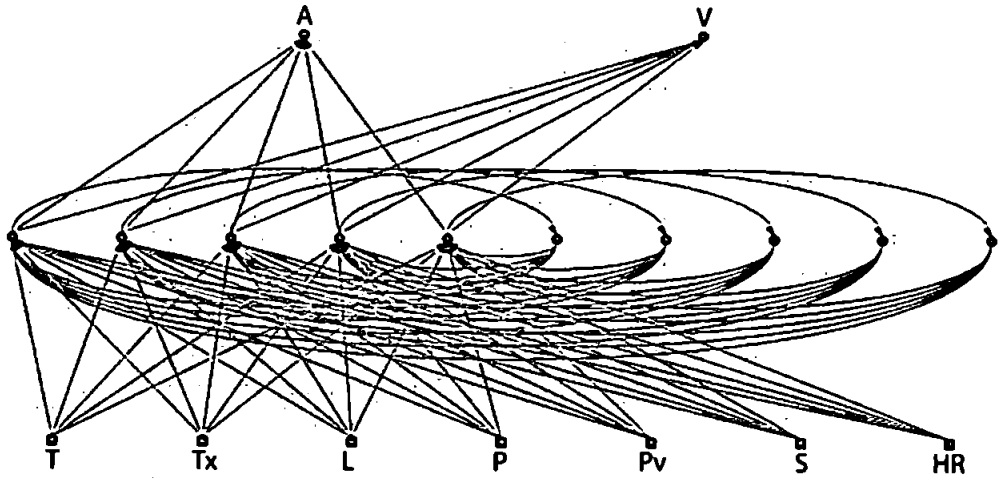


Figure 6.7: Neural network detailed architecture including all connections between processing units.

The lesioning tests consist of systematically "damaging" each of the connection weights in the model, and then testing it with the input data (all the music pieces). The output error quantifies how important this connection is to the model output dynamics. Figure 6.9 shows the resultant effect on the output predictions of removing each of the weights from the network. If the lesioned connection had little effect on the output performance ($rms < 0.090$), it is represented in black. This represents a low rms error and minimal contribution of this connection. Higher values of error are represented in grey ($0.090 < rms < 0.300$) and white ($rms >= 0.300$), and indicate the strength of influence of the weight.

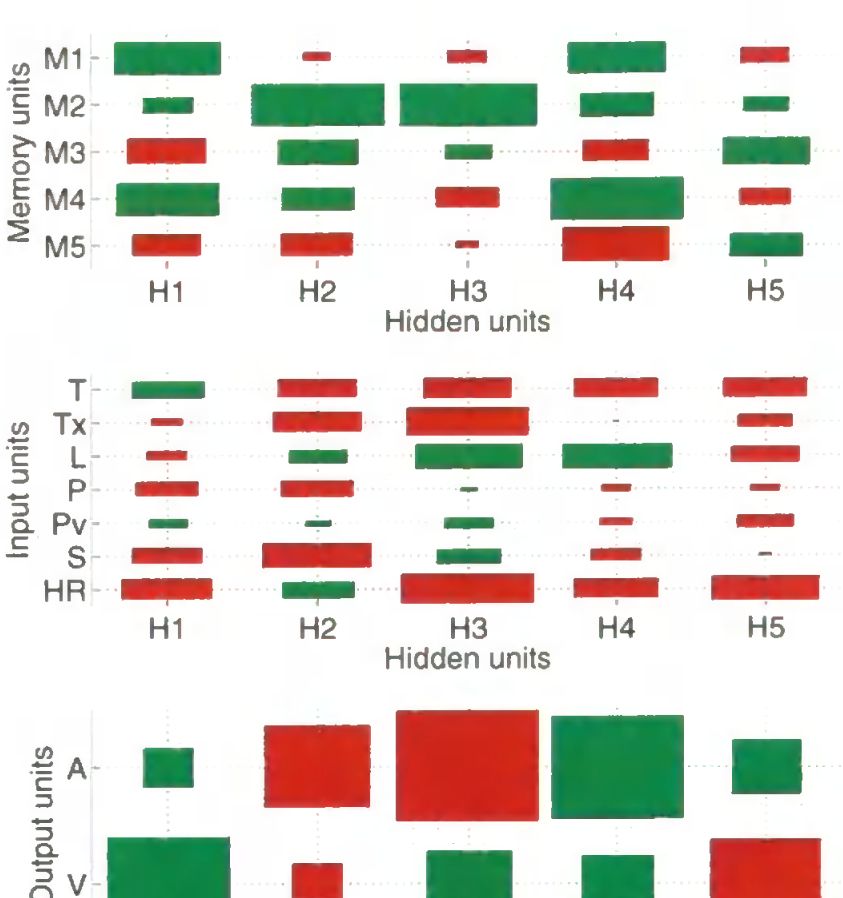


Figure 6.8: Neural network weight connection matrices: memory to hidden layers (top), input to hidden layers (middle) and hidden to output layers (bottom). The weights are represented as rectangles of variable size and colour: the size is proportional to the weight value and the colour represents the signal of the weight (red for negative and green for positive).

The lesioning tests show two distinct groups of hidden units with stronger effects on the output: H_2, H_3 and H_4 are related to arousal, while H_1, H_3, H_4 and H_5 are related to valence. All the sound features have significant impact on the model. At least two connections from each input to the hidden layer substantially reduced the model's performance (for arousal and valence).

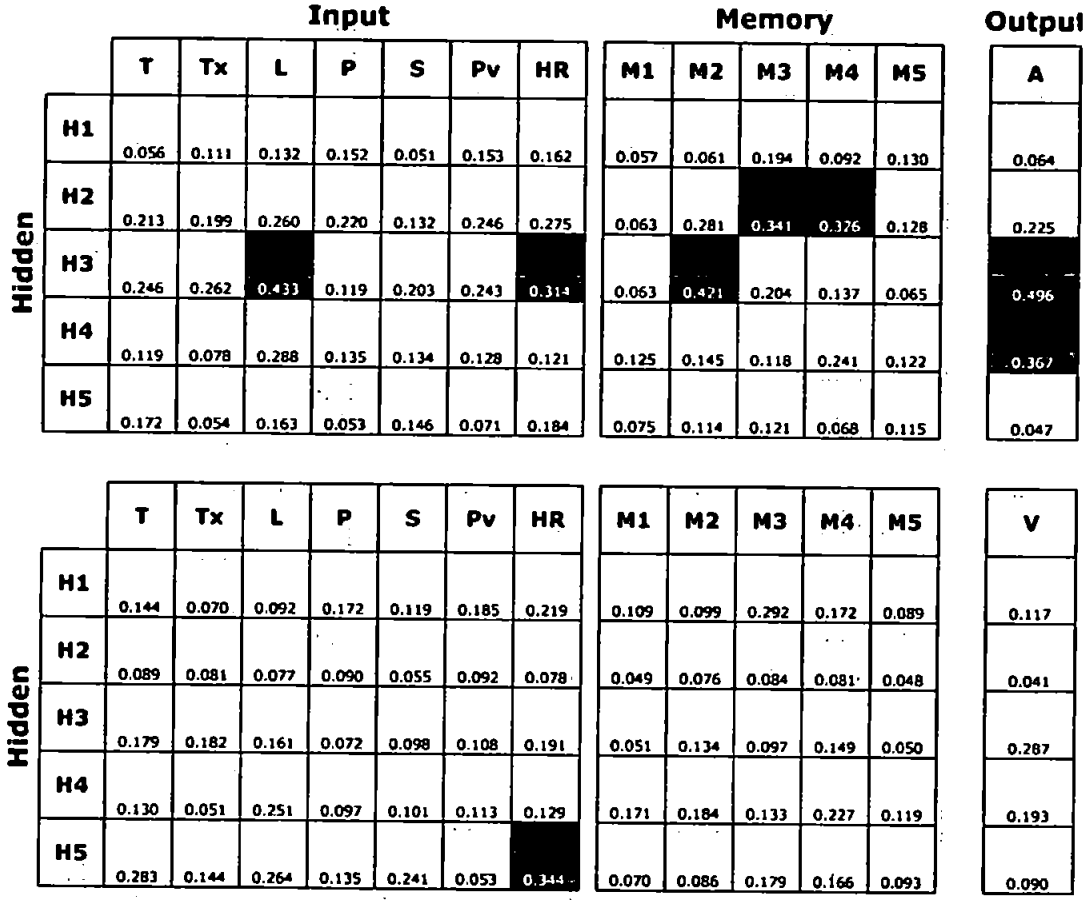


Figure 6.9: Model weight matrices analysis: each learned weight in the model was removed (value set to 0.0), one at a time. The model performance was measured using the rms error and the values are indicated for each weight removed. Each cell corresponds to the removal of one connection linking two processing units. For an easier reading the rms errors are represented using a colour code: black for those weights that had a small or no effects on the model performance ($rms < 0.09$); for higher errors between gray ($0.09 < rms < 0.30$) and white were used ($rms >= 0.30$).

As expected the temporal structure of the sound features were fundamental to the predictions of arousal and valence. With the exception of H_2, all the hidden units receive feedback from the memory units. These are the connections that "decode" the temporal structure of the inputs.

6.5.2 Model internal dynamics and input/output transformation

Due to the layered structure of the model, the analysis of the input/output transformation rules requires an analysis of the input/hidden and hidden/output transformations separately. The activity of the hidden units is also determined by the memory units activity. This makes it difficult to identify the precise contribution of each input to the output activity. In order to reveal some of those relationships, I will analyse the correlations between the inputs and hidden units activity. In this way, it is possible to investigate the input information is passed to the outputs.

The canonical correlation analysis (CCA) (Hotelling, 1936) was used again to investigate the dynamics of data flow within the model (see Table 6.11).

Two canonical variables explain 91.6% of the variance in the data. The first pair of variables loads on L, T, P, S, HR (input set), H_3 and H_5 (hidden layer). The second, loads mainly on T and H_1.

By considering the weights from the hidden units to the output it is possible to establish qualitative patterns of correlations illustrative of the general model dynamics. The lesioning tests allowed selecting the groups of hidden units with direct relationships to each output, while the CCA showed how the hidden representations vary with the inputs. By considering both analyses together it is possible to symbolically represent the relationships between the inputs and the outputs[2].

The first canonical function loads positively on L, T, P, S and HR, and negatively on H_3 and H_5. H_3 has its strongest impact on the arousal output with a

[2]The signal of the canonical loadings indicates if a hidden unit reinforces or inhibits the outputs.

Loadings (Input/Hidden)		
Variable	var. 1	var. 2
H_1	0.046	**-0.555**
H_2	-0.484	0.302
H_3	**-0.779**	0.425
H_4	-0.037	0.338
H_5	**-0.899**	-0.347
L	**0.421**	-0.312
T	**0.494**	**-0.780**
P	**0.352**	-0.071
Pv	0.033	0.085
Tx	-0.209	0.114
S	**0.430**	-0.284
HR	**0.892**	0.416
Canon Cor.	0.976	0.895
Pct.	76.5%	15.1%

Table 6.11: Canonical Correlation Analysis (CCA): the canonical correlations (the canonical correlations are interpreted in the same way as the Pearson's linear correlation coefficient) quantify the strength of relationships between the extracted canonical variates, and so the significance of the relationship. To assess the relationship between the original variables (inputs and hidden units activity) and the canonical variables, the canonical loadings (the correlations between the canonical variates and the variables in each set) are also included.

negative weight ($w_{H3-A} = -3.78$), while H_5 has a negatively weighted connection to the valence output ($w_{H5-V} = -2.27$). In this way those inputs have a general positive effect on both outputs. The second canonical function consists mainly of T and H_1. Because the canonical function loads negatively on T and H_1, which in turn has a positive weight to the output ($w_{H1-V} = 2.720$), this dimension conveys a positive effect of T on valence. These relationships are coherent with the analysis conducted in Chapter 4, with the addition of the positive effect of heart rate on the arousal and valence outputs.

Pv and Tx are the only variables with no significant linear correlations with the hidden layer. This does not mean that they are not relevant for the model dynamics, but instead that they have more complex interactions with the output. As seen in the lesioning analysis (see Section 6.5.1) all the inputs

give an important contribution to the model. Unfortunately these relationships are not necessarily linear and exclusive, making it difficult to analyse the model performance. This fact also suggests that the model representations are distributed, meaning that the hidden units distribute the input signals through the internal representations space. This is a strong argument to suggest that interactions among the sound features are a fundamental aspect of the affective value conveyed.

6.6 Discussion

In this Chapter, I presented a series of simulation experiments on two new models of emotional responses to music. Model 1 solely used sound features as inputs. and the self report data obtained in the experiment (described in the previous chapter) as the outputs. This model has reproduced the work presented in Chapter 4. The second model, Model 2, is a modification of Model 1 that includes physiological variables as inputs for the model. The objective was to evaluate the possible contribution of peripheral feedback to the subjective feeling component of emotion.

The initial tests on the model architecture, using the new experimental data, showed that the best model configuration was similar to the one developed in Chapter 4 (Model 0). Tempo, loudness, sharpness, pitch variation, mean pitch and texture are the sound features which in both models permitted "mimicry" of human subjective feelings for a set of pieces, and predicted new ones for another set. These results support the hypothesis that the psychoacoustic properties of sound contain relevant information about the affective experience with music. Moreover, spatiotemporal neural networks can simulate those relationships.

The results obtained with Model 1 showed a comparative improvement to Model 0, especially in terms of valence predictions (the linear correlation between

the model's outputs and human participants responses for arousal were similar, but for valence it improved by approximately 8%). Such an improvement in performance can either be attributed to the fact that the sound features used in both models (they are the same) are better at describing the pieces used in Model 1. Alternatively it is possible that the change in emphasis in the experiment (the participants were asked to report the emotion "felt", rather than the one "thought to be expressed" by the music) is responsible. This work does not enable to identify with certainty the reasons why this happened, nevertheless previous research suggests that both reasons may be partially responsible for this result (Zentner et al., 2000). A possible solution to this ambiguity would be to investigate this issue, including all the pieces used for both models, with two groups of participants: a group reporting the emotion "felt", and another the emotion thought to be "expressed" by the music. Then, the results could be compared.

Model 2 was essentially an experiment to test the peripheral feedback hypothesis. The model was extended to include the physiological variables as inputs, together with the sound features. In agreement with the analysis of the experimental data in Chapter 5, the inclusion of the heart rate level added insight to the subjective feeling variations. The inclusion of this physiological variable produced an improvement to the subjective feeling predictions: the correlation between the model outputs and human participants' responses of 10% for arousal and almost 20% for valence. These results suggest that the heart rate variations may be used as an insight to the emotion "felt" while listening to music, as demonstrated by Dibben (2004). A detailed analysis of the model has also shown that the heart rate has a positive relationship with reports of arousal and valence.

Figure 6.10 provides a qualitative representation of the relationships between sound features and affective dimensions as in Chapter 4, updated with the

relationship of the heart rate level, to arousal and valence. Note that this representation is representative of the main observable fluxes of information in the model, and does not represent the complete interaction between inputs and outputs.

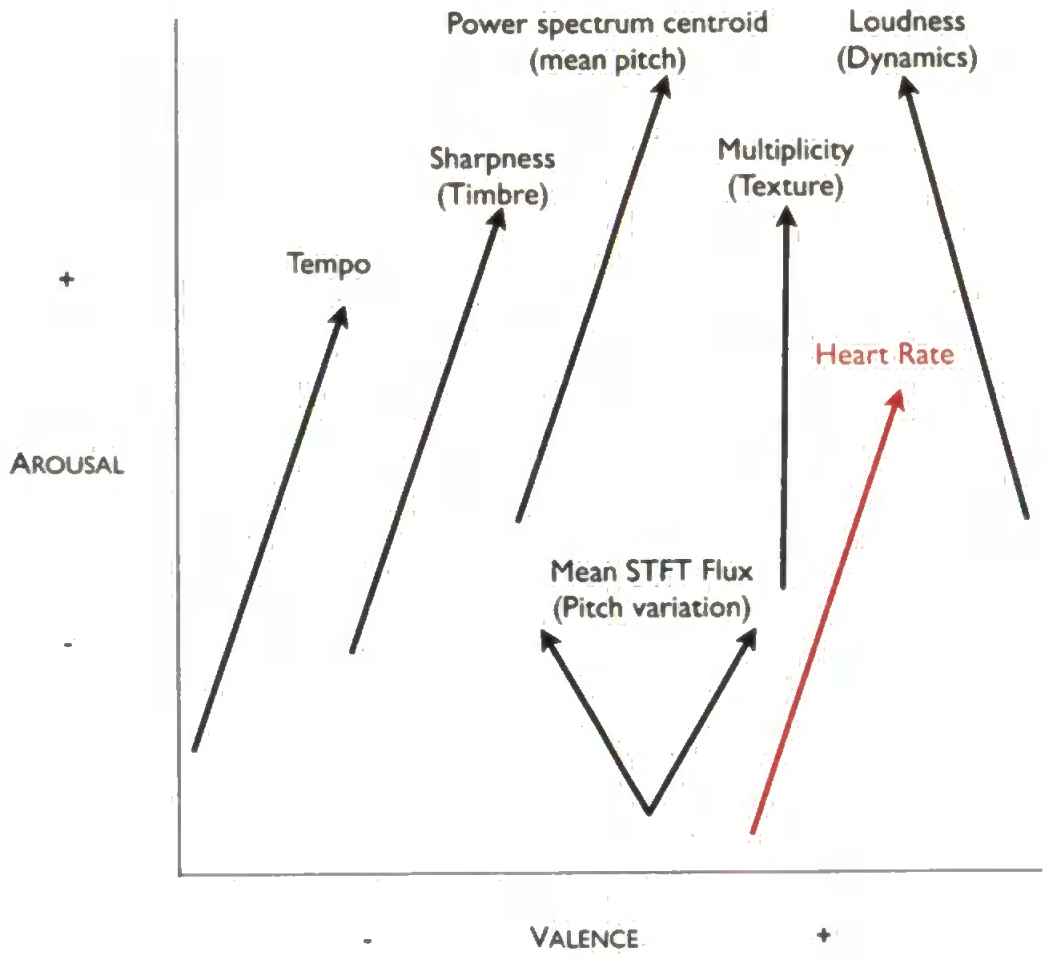


Figure 6.10: Qualitative representation of individual relationships between music variables, hear rate and emotion appraisals: summary of observations from model analysis. The direction of the arrows indicates an increase in the variable indicated (the arrow sizes and angles formed with both axis are merely qualitative, and cannot be interpreted in mathematical terms)

As you can see most variables have a positive effect on both arousal and valence. The exceptions are loudness, which tends to have a negative effect on valence when it increases, and the mean STFT flux (pitch variation), that shows

more complex relationships with valence (having both positive and negative effects).

Chapter 7

Conclusions and future research

7.1 Summary and conclusions

The central focus of this dissertation was the development of a computational model, combined with experimental investigations, capable of predicting human participants' self report of subjective feelings of emotion, while listening to music. It was hypothesised that the affective value of music may be derived from the nature of the stimulus itself: the spatiotemporal patterns of sound (psychoacoustic) features that the brain extracts while listening to music. In this way, the link between music and emotional experiences was considered to be partially derived from the representation of the musical stimulus in terms of its affective value. The belief is that organised sound may contain dynamic features, which may "mimic" certain features of emotional states. For this reason this investigation was based on continuous representations of both emotion and music.

In Chapter 2, the theoretical and experimental background for this thesis was presented. Emotions were discussed in a neurobiological context, emphasising brain states and bodily responses as the substrate of emotions and determinant aspects of conscious feeling states. By analysing the different mechanisms

by which the construct of emotion may arise or manifest itself, I focused on physiological cues and subjective feelings as two important qualities to study musical emotions. The quantification of these components was also addressed in order to establish an experimental framework that allowed me to focus on continuous measurements of emotion, since music is characterised by constant changes over time. The intention was to focus on emotion itself and not on mood states. The time scale in emotion studies is of importance, especially because musical emotions may exhibit time locking to variations in psychological and physiological processes (e.g. Goldstein, 1980; Nielsen, 1987; Krumhansl, 1997; Schubert, 1999a; Korhonen, 2004a).

An aspect that became evident in Chapter 2 was 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. Chapter 3 addresses this question by considering a novel methodology for their study. The Elman neural network, a class of spatiotemporal connectionist models, was suggested as a paradigm capable of analysing the interaction between sound features and the dynamics of emotional ratings. The initial evaluation of this modelling framework was described in Chapter 4.

Chapter 4 described the first novel investigation. It presents an investigation of musical emotions based on an Elman neural network, a modelling paradigm that supported the investigation of both temporal dimensions (the dynamics of musical sequences) and spatial components (the parallel contribution of various psychoacoustic factors) to predict human participants' emotional ratings of various music pieces. The network was trained on the subjective feelings data (valence and arousal ratings) collected in a previous study by Korhonen (2004a). After training, the network was able to successfully generalise the emotional ratings of music, including novel musical sequences from a set of six key sound features: loudness (dynamics), power spectrum centroid (mean pitch),

sharpness (timbre), mean spectral flux (pitch variation), multiplicity (texture) and tempo.

In terms of the modelling technique, the computational model presented constitutes an advance in three essential aspects. First, it incorporates all music variables together in a single model. This is an important aspect as an acknowledged limitation of studies into the relationships between music features and emotion is the fact that different variables are usually considered in isolation (Gabrielsson & Lindström, 2001). This is an oversimplification of the relationships between sound features and emotional response, and it creates problems at the analysis level. As a consequence, it leads many times to contradictory "rules systems". In this work some of the drawbacks of analysing time series of music and emotion were overcame by considering the interactions between sound features. The second aspect is that the use of neural networks, as nonlinear models, permits a richer representation of the spatiotemporal affective features of music. These were not accessible to linear modelling techniques such as the ones used by Schubert (1999a) and Korhonen (2004a). The model used in Chapter 4 can also operate in higher dimensional spaces (this fact is especially important for valence predictions, which have shown to be more complex relationships with sound features). The third relevant aspect of the model presented was the analysis of its internal feature space (the spatiotemporal representations of the affective value of the encoded music), which has shown meaningful relationships between sound features and emotion, as summarised in Figure 6.10. In my opinion these are strong arguments to support the use of spatiotemporal connectionist models in the study of musical emotions inline with the discussion presented in Chapter 3.

Another positive aspect of the model presented was its performance on novel music. The neural network was trained using 486 seconds of music (3 pieces). These were the "sample" pieces for the model, used to make it

"mimic" as closely as possible to the human participants. Then, three new pieces (other 632 seconds of music), unknown to the model, were used to test it. It was shown that the model predictions also resemble those obtained from human participants. There is another important argument suggesting the preponderance of spatiotemporal structures of sound (organised as music) and emotion dynamics (in relation to subjective feeling states). As referred in Chapter 3, a computational model, being an abstract representation of the system under study, can be used as a platform to study some aspects of its nature. That was the principle that led to a new investigation.

In Chapter 5 a new experimental study was presented which used the continuous response methodology to obtain the listener's affective experience with music. It was hypothesised that music can alter both the physiological component and the subjective feeling of emotion in response to music, sometimes in highly synchronised ways. In order to do so, the subjective feelings reported by human participants while listening to music were obtained, together with the participants' heart rate and skin conductance levels. The analysis of the experiment has shown that loudness, tempo, mean pitch, sharpness and timbral width have a positive effect on both psychological arousal and valence (a stronger effect was observed on the first). Regarding the physiological recordings, only the heart rate showed statistically significant relationships to the arousal dimension of emotion. Further analysis of the synchronisation between "peak" changes in the physiological and subjective feeling reports confirmed that result and revealed another: almost half of the strong changes in heart rate were followed by changes in arousal and valence - 43% of the strong changes in heart rate preceded (up to 4s) strong changes in the subjective feeling component.

Evidence provided in the Chapter 5 also suggested the existence of relevant interactions between the psychological and physiological components of emotion. A higher differentiation on the general levels of arousal and valence was obtained

when using the physiological dimensions together with psychoacoustic variables, when compared with only psychoacoustic variables. In 81.5% of the test cases, combinations of the sound features resulted in a successful classification of the general affective value (2DES quadrant) of each segment. By combining sound features and physiological variables, the rate increased to 92.6%. This improvement suggests that physiological cues combined with sound features, give a better description of the self report dynamics. As discussed, those results are consistent with previous findings reported in the literature (see Section 5.4).

The data that was collected empirically was then used in Chapter 6 in a new computational investigation: the extension of the neural network model presented in Chapter 4, to include physiological cues. The first aim of this investigation was to reproduce the model developed with Korhonen's data. The model developed in Chapter 6, which included only the sound features as input ("Model 1"), had the exact same architecture as the model presented in Chapter 4 (Model 0). This is another convincing result that supports the idea that listeners may derive affective meaning from the spatiotemporal patterns of sound, reinforced by the fact that the model was tested on a new population and on a different set of music pieces.

Overall, the studies presented in this thesis provide a computational framework capable of inferring the affective value of music from structural features of the perceptual (auditory) experience. The neural network model used focused in detecting temporal structure in the perceptual dimensions of loudness, tempo, texture, timbre and pitch, rather than only on delays between music events and affective responses, as in previous studies (Schubert, 1999a; Korhonen, 2004a). Moreover, it integrates all variables in a single computational model, able to detect structure in sound - either as interactions among its dimensions or their temporal behaviour - and to predict affective responses from its form. Along with the development of the model, there was a strong emphasis on the generalisation for novel music. A good generalisation performance indicates a

potentially valid model and a good platform to analyse the relationships between sound features and affective responses (as encoded by the model). At the analysis level, modelling and experimental results, provide evidence suggesting that spatiotemporal patterns of sound resonate with affective features underlying judgements of subjective feelings. A significant part of the listener's affective response is predicted from the a set of six psychoacoustic features of sound - tempo, loudness, multiplicity (texture), power spectrum centroid (mean pitch), sharpness (timbre) and mean STFT flux (pitch variation).

The final investigation regarding the peripheral feedback hypothesis has shown that the inclusion of the heart rate level as an input to the model (a source of information to predict the affective response) improves the model performance. The correlation between the model predictions and human participants' responses, was improved by 10% for arousal and almost 20% for valence. This is a supporting argument favouring the peripheral feedback theories, emphasised by the fact that strong changes in the heart rate level consistently preceded changes in listeners' self report of arousal.

7.2 Contribution to knowledge

This thesis contributed some evidence that may help to clarify certain aspects of emotional reactions to music. I have presented strong evidence suggesting that spatiotemporal patterns of sound may resonate with certain affective features underlying judgements of subjective feelings. The same model architecture was applied to data from two different experiments (including two different groups of participants and music pieces) and in both cases the model was able to predict human participants' responses, reinforcing the importance of these findings.

Reinforced by the model performance and resemblance to other empirical work in terms of the relationships between music features, arousal and valence,

the effect of peripheral feedback in the subjective feelings reports was also evaluated. Following evidence presented in other studies (Dibben, 2004; Philippot et al., 2002), I tested how physiological cues affect the perception of emotion in music. Results suggest that at least the heart rate variations may serve as a clue to infer the intensity and quality of the emotion "felt" while listening to music. A relationship of causality between physiological events and reevaluations of "felt" arousal and valence was also found, which reinforces the argument in towards the peripheral feedback hypothesis.

More generally this thesis has presented some evidence supporting the "emotivist" views on musical emotions. It was shown that a significant part of the listener's affective response (at least to classical music) can be predicted from the psychoacoustic properties of sound. Those sound features, to which Meyer referred as "secondary" or "statistical" parameters, encode significant parts of the information that enables the approximation of human affective responses to music. Contrary to Meyer's (1956) belief, the results in this thesis suggest that "primary" parameters, derived from the organisation of secondary parameters into higher order relationships with syntactic structure, may not be a necessary condition for emotion to arise. This result is also coherent with the study of Peretz et al. (1998), in which a patient lacking the cognitive capabilities to process the music structure (including Meyer's "primary" parameters), was able to identify the emotional tone of music.

Finally, this thesis provides a methodological contribution to the field of music and emotion research. I developed a novel methodological approach, 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.

7.3 Future research

The findings presented throughout this dissertation have several implications for the study of emotion in music. Nevertheless, this is only a step forward since there are some limitations that still need to be addressed. Moreover, new tests are needed to assess the potential of this model.

Experiments The music pieces used in this thesis did not cover the whole areas of the possible affective response. Further work should address this issue, for instance, by including in the experimental methodology a pretest of affective value of potential pieces (e.g. adjective checklists).

Regarding the physiological measurements, skin conductance response did not show contextual relationships with affective events. Instead, heart rate was consistently found related to both sound features (tempo) and self report of affective response (arousal and valence). It is possible that other physiological indexes contain relevant features of affective experiences. An interesting extension of the physiological variables set would be to investigate respiration related variables.

The number of participants to the experiments should also be increased. It is interesting to observe the result to larger groups of listeners, particularly because it permits to analyse separate groups of listeners, group accordingly to certain categories (e.g. based on gender, age, musical expertise, personality tests). Also interesting would be to operate a change in the experimental context, shifting from experiment rooms to concert halls (or multimedia rooms). That would allow the synchronous measurement of affective responses to music of groups of people. Additionally, it would reduce substantially the time consumption of the experiment, and investigate other realistic listening scenarios.

Computational model The computational model presented was applied to a limited set of music and listeners (western instrumental/art music). The focus of this thesis on classical music is due to the fact that this musical style is the most often studied in music and emotion research. Another reason is that classical music is widely considered to be a style which "embodies" emotional meaning (or better, in which the expression and perception of emotion are closely related). A fundamental aspect that needs to be addressed in future research is certainly the expansion of the musical universe. I would suggest that the initial studies should be done separately for different styles and subsequently they should be merged. It would then be possible to analyse the extent to which the model presented might be applied to a broader range of music styles.

Due to the flexibility of connectionist models it is also possible to extend the model presented to incorporate other variables related to perceptual (e.g. psychoacoustic features or higher order parameters such as harmony or mode), physiological (e.g. respiration measures) or even individual characteristics of the listener (e.g. musical training/expertise, personality traits, among others). The modelling process may also be improved by using other types of spatiotemporal connectionist models. Another class of models to consider are the long short-term memory neural networks (Hochreiter & Schmidhuber, 1997). These networks are able to represent more complex temporal relationships of the modelled processes, have a good generalisation performance and the ability to bridge long time lags. This characteristic may be relevant when considering the effects of music on "mood" states or the accumulated effects of specific variables (sound features or physiological variables).

Another important aspect which deserves further attention is the visualisation of the model dynamics. The complex relationships between the sound features and the subjective feeling component of emotion created difficulties in identifying the model transformation rules, especially for the sound features related to pitch

variation and texture. One possible solution to this problem might be to consider the use of more sophisticated data visualisation techniques (e.g. Stuart, Marocco, & Cangelosi, 2005) or even a precise mathematical description of the model.

Applications Finally, in addition to the music specific considerations, the analysis of affective cues in speech would be an interesting experiment. There is evidence that vocal expressions can be universally recognised through emotion specific patterns of voice cues (e.g. Laukka, 2004; Juslin & Laukka, 2003). Supported by empirical evidence, they have suggested that music performance involves similar patterns of acoustic cues conveying affective meaning similar to those of speech.

By considering the acoustic patterns of speech and music in the same model it could be possible to study their relationships. If the model can predict the emotional qualities of both speech and music cues, then it would be an important argument in favour of the existence of a cross-cultural musical expression (similar to vocal and facial expressions). For this work it would be interesting to study preverbal and verbal societies in activities involving speech, music or both.

Health care and music therapy are potential areas of application for the model. It is possible to develop models of single and groups of listeners, as an auxiliary tool for the analysis and improvement of sound spaces in health care institutions or public spaces. It can also be developed towards other purposes, such as the prediction of affective value of music pieces as a therapeutic tool. It can also be used as a component of a computer music system, and used as a tool composing new music.

Appendix A

Experiment: call for participants

The experiment was publicised using printed and electronic formats. The information was distributed inside and outside the University of Plymouth using flyers, emails and posts to electronic portals. The contents of all the formats were the same and are shown below (flyer format):

Study: Music and Emotions

The University of Plymouth is running an experiment on Music and Emotions. Participants will be paid £6 for a 1h study, and will be required to listen to music and report their emotional state.

£6 for 1 hour

The only requirement is that participants appreciate listening to music, especially western classical music.

Study running: July and August

If you are interested in taking part in this study, please email eduardo.coutinho@plymouth.ac.uk.
Alternatively you can call 01752232548 or 07910111447.
You also can see us in Portland Square Building (room B110) at the University of Plymouth. The experiment will take place in room A216 (Portland Square Building)

Figure A.1: Call for participants: this information was distributed in printed and electronic formats.

Appendix B

Questionnaire

At the beginning of the experiment, participants answered a questionnaire regarding their musical education and experience, exposure to and enjoyment of classical music. In the questionnaires participants were asked for the frequency of exposure and enjoyment of classical music, and their level of music training/education (see below). Participants used using 5-point Likert scales to answer to these questions. At the end of each piece the participants were asked to rate how much they liked the piece they have just listened to, as well as they familiarity with it.

Participant nr. ___

Gender (M/F) _____ Right handed ☐ Left handed ☐

Age _____

Mother tongue _____

Nationality _____

Occupation _____

1. How many years of traning do you have on a musical Instrument (or in singing)?

0-1 years ☐ 1 1-3 years ☐ 2 3-5 years ☐ 3 5-10 years ☐ 4 10+ years ☐ 5

2. How much exposure do you have to Western instrumental art / classical music? *(rate from 1 to 5)*

(none) ☐ 1 (rarely) ☐ 2 (ocasionally) ☐ 3 (frequently) ☐ 4 (a lot) ☐ 5

3. How much do you enjoy listening to Western instrumental art / classical music? *(rate from 1 to 5)*

(I hate) ☐ 1 (I don't like) ☐ 2 (Neutral) ☐ 3 (I like) ☐ 4 (I love) ☐ 5

Music 1

How much did you like this music? (rate from 1 to 5)

(I hated) ☐ 1 (I didn't like) ☐ 2 (Neutral) ☐ 3 (I liked) ☐ 4 (I loved) ☐ 5

Are you familiar with this piece?

Music 2

How much did you like this music? (rate from 1 to 5)

(I hated) ☐ 1 (I didn't like) ☐ 2 (Neutral) ☐ 3 (I liked) ☐ 4 (I loved) ☐ 5

Are you familiar with this piece?

Music 3

How much did you like this music? (rate from 1 to 5)

(I hated) ☐ 1 (I didn't like) ☐ 2 (Neutral) ☐ 3 (I liked) ☐ 4 (I loved) ☐ 5

Are you familiar with this piece?

Music 4

How much did you like this music? (rate from 1 to 5)

(I hated) ☐ 1 (I didn't like) ☐ 2 (Neutral) ☐ 3 (I liked) ☐ 4 (I loved) ☐ 5

Are you familiar with this piece?

Music 5

How much did you like this music? (rate from 1 to 5)

(I hated) ☐ 1 (I didn't like) ☐ 2 (Neutral) ☐ 3 (I liked) ☐ 4 (I loved) ☐ 5

Are you familiar with this piece?

Music 6

How much did you like this music? (rate from 1 to 5)

(I hated) ☐ 1 (I didn't like) ☐ 2 (Neutral) ☐ 3 (I liked) ☐ 4 (I loved) ☐ 5

Are you familiar with this piece?

Music 7

How much did you like this music? (rate from 1 to 5)

(I hated) ☐ 1 (I didn't like) ☐ 2 (Neutral) ☐ 3 (I liked) ☐ 4 (I loved) ☐ 5

Are you familiar with this piece?

Music 8

How much did you like this music? (rate from 1 to 5)

(I hated) ☐ 1 (I didn't like) ☐ 2 (Neutral) ☐ 3 (I liked) ☐ 4 (I loved) ☐ 5

Are you familiar with this piece?

Music 9

How much did you like this music? (rate from 1 to 5)

(I hated) ☐ 1 (I didn't like) ☐ 2 (Neutral) ☐ 3 (I liked) ☐ 4 (I loved) ☐ 5

Are you familiar with this piece?

OBSERVATIONS:

Appendix C

Participants Information Sheet

Before the experiment started participants were given a written description of the experiment as copied below:

"You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. This study aims to obtain your personal emotional experience with the music you are going to listen to. To do so we will ask you to listen to 6 pieces of music and to report your emotional experience continuously in a computer framework by using a mouse. Together with it we will measure some of your physiological reactions during listening. This data will be used to analyse both your psychological and physiological reactions and compare it with the psychoacoustic properties of the music. You will be asked to fill an additional questionnaire that aims to collect relevant personal information related to your appreciation of each music, the music style and also some personal details (age and mother tongue). We will ask you to complete another questionnaire to obtain more information about you personality. The questionnaires information will be used to classify you in a specific group in order to better interpret the experimental data obtained. This

is part of our hypothesis. The experiment will take approximately 1 hour. First youll be familiarised with the self report framework by looking at several pictures and identify their emotional content. This is a standard test and its only aim is to make you comfortable in using the self report framework. Then youll listen to each piece of music and during the process youll be asked to rate your emotional reaction to it in a continuous time scale. At the end of each piece youll be also asked to say how much you enjoyed the music and also if you knew the piece. At the same time, for all pieces, we will be recording your physiological activity. The final questionnaire will be filled at the end of the listening tasks and aims to obtain some information about your personality, namely the direction of energy expression, the method of information perception by a person, how the person processes information, and how a person implements the information he or she has processed. We have chosen you to participate in this experiment because we are searching for a heterogeneous group of people. The total number of participants will be 60. Your participation can help us to better understand the process of emotion during music listening which can help to convey information about the role of music in human life and therefore be applied in several areas such psychology research and music therapy.

All the information obtained from you during this study will be kept strictly confidential. Your data will not be associated with your name at any time, and only the Researcher conducting the study will have access to it. If it is used for a future publication the average data of the experiments will be used and so no Participants individual details will be given at any time. The final results of this study will be used for the Researchers PhD thesis and youll be able to find them in www.eadward.org.

The Researcher is conducting the research as a student at the University of Plymouth, in the School of Computing, Communications and Electronics. This research is funded by the Portuguese Foundation for Science and Technology

(FCT, Portugal).

This research has been approved by the Faculty of Technology Ethics Committee.

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. Thank you for taking the time to read this information sheet."

Appendix D

IAPS pictures

Picture	IAPS code	Arousal/Valence
Rafting scene	8370	high/high
Scene of violence	3530	high/low
Tiny rabbit	1610	low/high
Slit throat	3071	high/low
Graveyard	9220	low/low
Basket	7010	neutral/neutral
Scene in a hospital	2205	low/low
Erotic female	4240	high/high
Spoon	7004	neutral/neutral
Child	9070	low/high

Table D.1: Pictures used in the experiment to test participants understanding of of the 2-dimensional affective space of Arousal and Valence. The pictures in the table were obtained from the International Affective Picture System (IAPS) database.

Appendix E

Randomised order of pieces presentation

Piece/Order	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
Music 1	5	4	5	6	5	3	10	4	3
Music 2	3	9	4	6	9	4	3	2	5
Music 3	4	2	10	4	3	7	6	7	2
Music 4	7	3	5	3	5	7	3	5	7
Music 5	10	6	3	3	5	2	3	5	8
Music 6	2	4	4	6	5	7	6	4	7
Music 7	5	5	5	5	4	4	4	10	3
Music 8	6	6	3	7	4	5	5	4	5
Music 9	3	6	6	5	5	6	5	4	5

Table E.1: Number of times each music was played in which order during the experiment.

Appendix F

Sound features visualisation

Figures F.1 to F.9 show the sound (psychoacoustic) features extracted using *PsySound 3* (2008) and used in Chapters 5 and 6.

Figure F.1: Sound features extracted with Psysound 3: T. Albinoni - Adagio (piece 1)

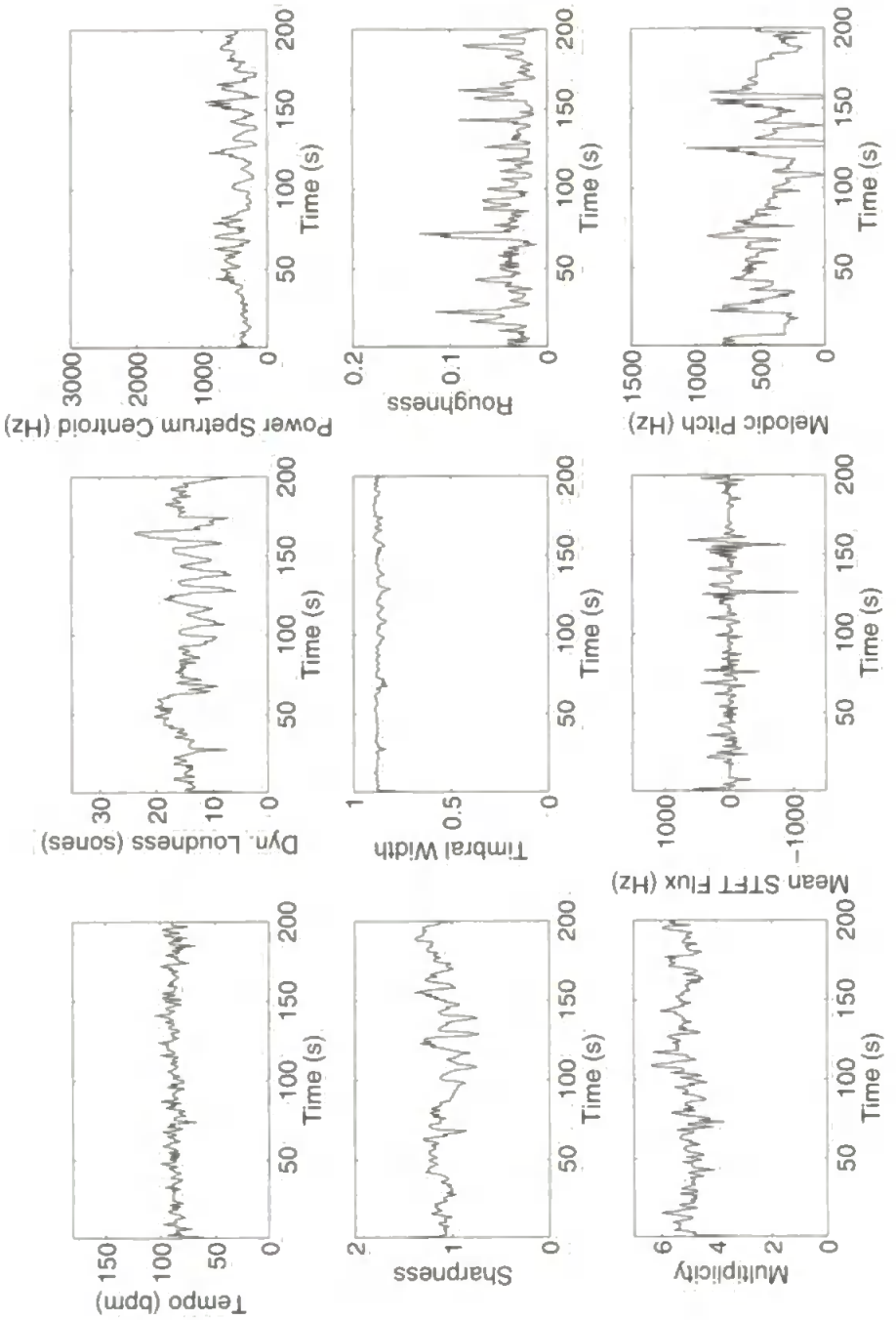


Figure F.2: Sound features extracted with Psysound 3: E. Grieg - Peer Gynt Suite No. 1 - IV. "In the Hall of the Mountain King" (piece 2)

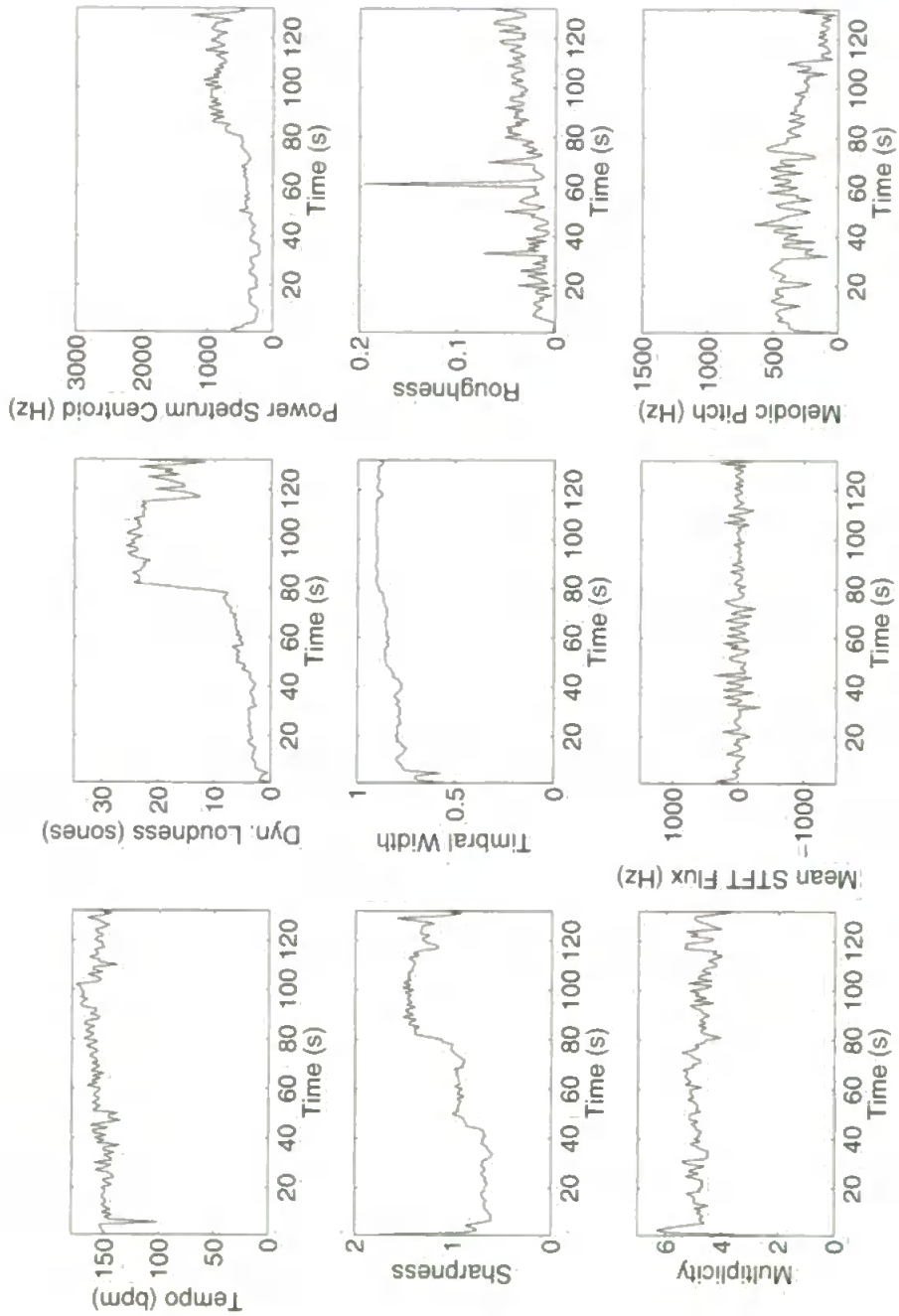


Figure F.3: Sound features extracted with Psysound 3: J. S. Bach - Prelude and Fugue No. 15 - I. "Prelude" (piece 3)

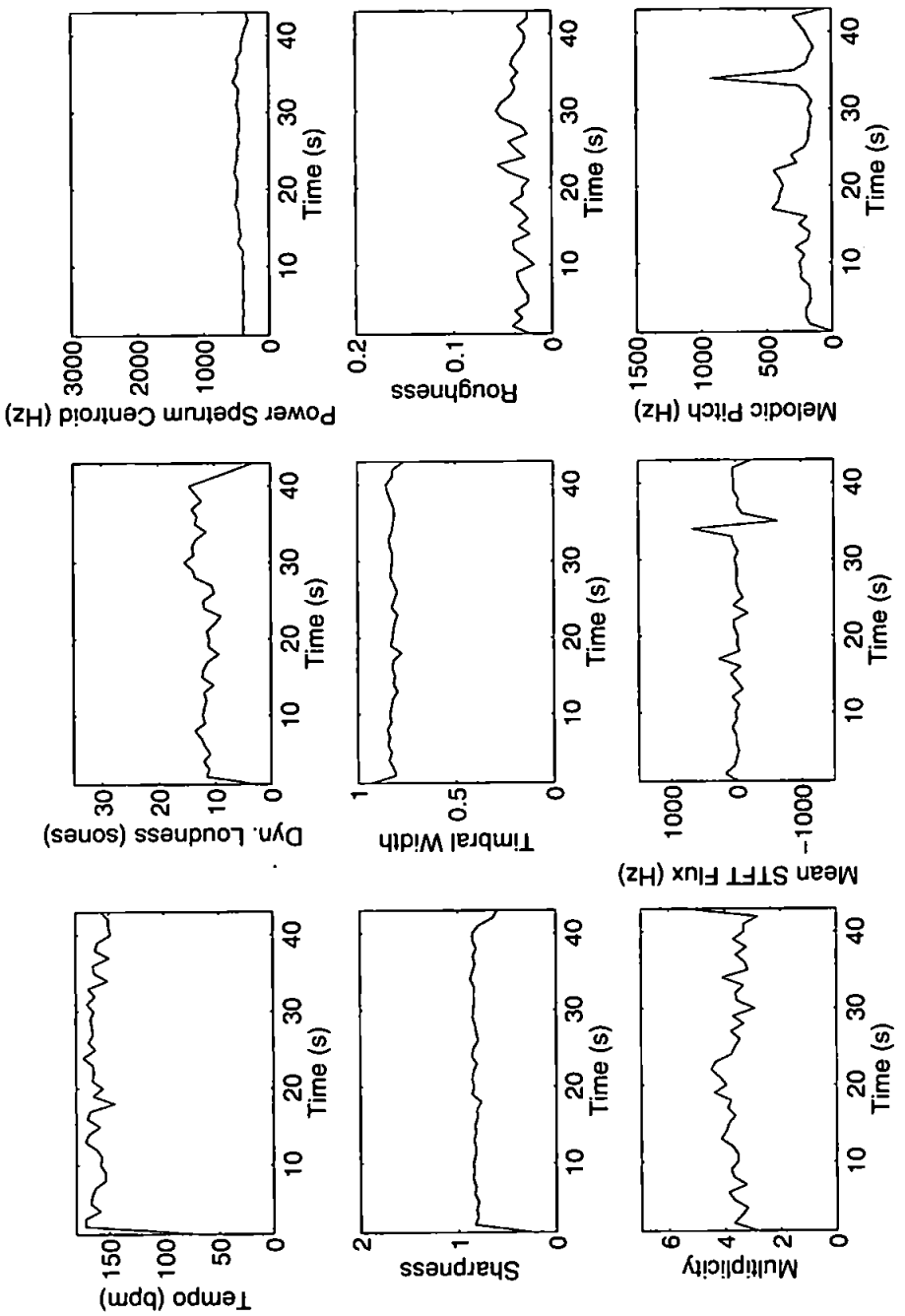


Figure F.4: Sound features extracted with Psysound 3: L. V. Beethoven - Romance No. 2 (piece 4)

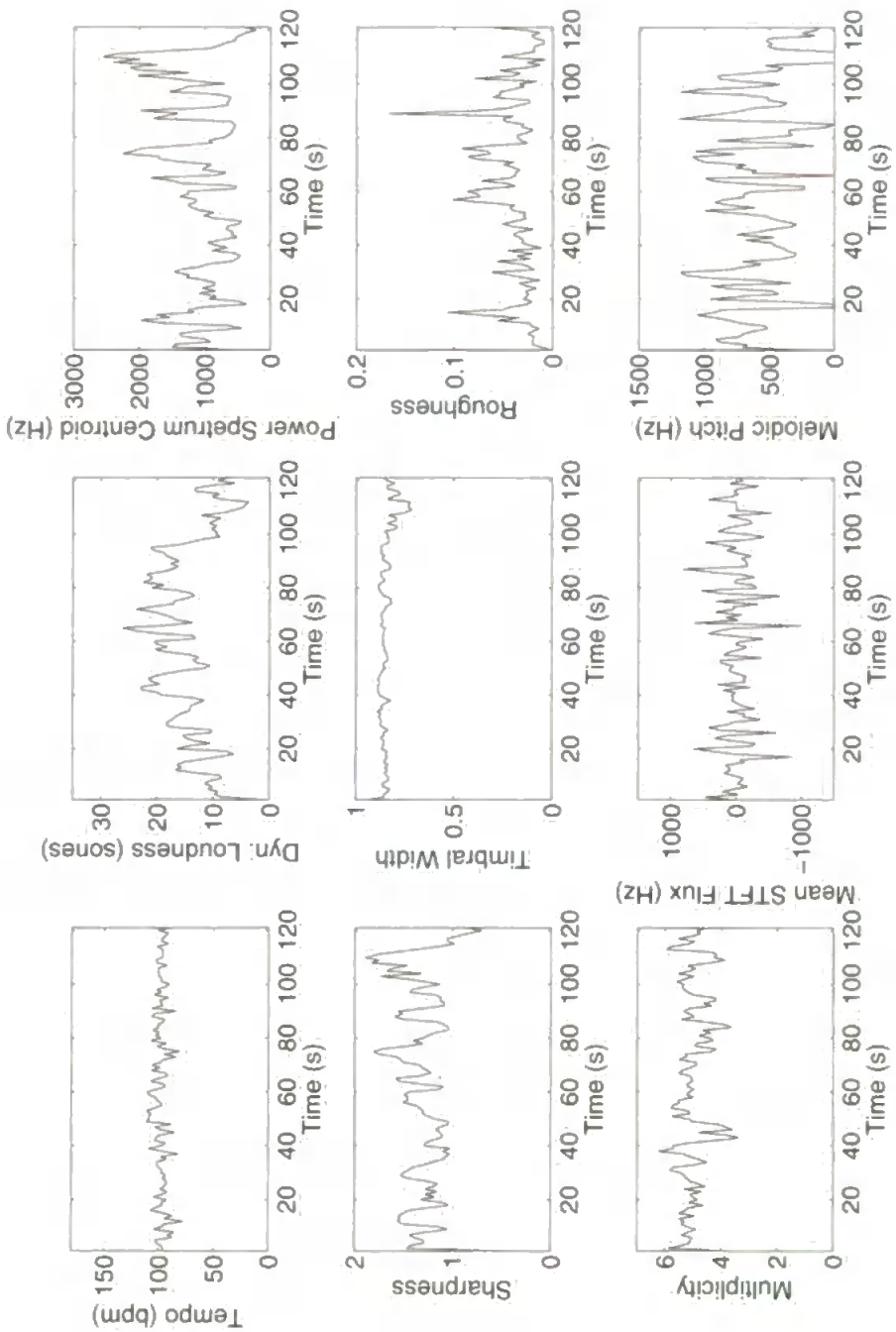


Figure F.5: Sound features extracted with Psysound 3: F. Chopin - Nocturne No. 2 (piece 5)

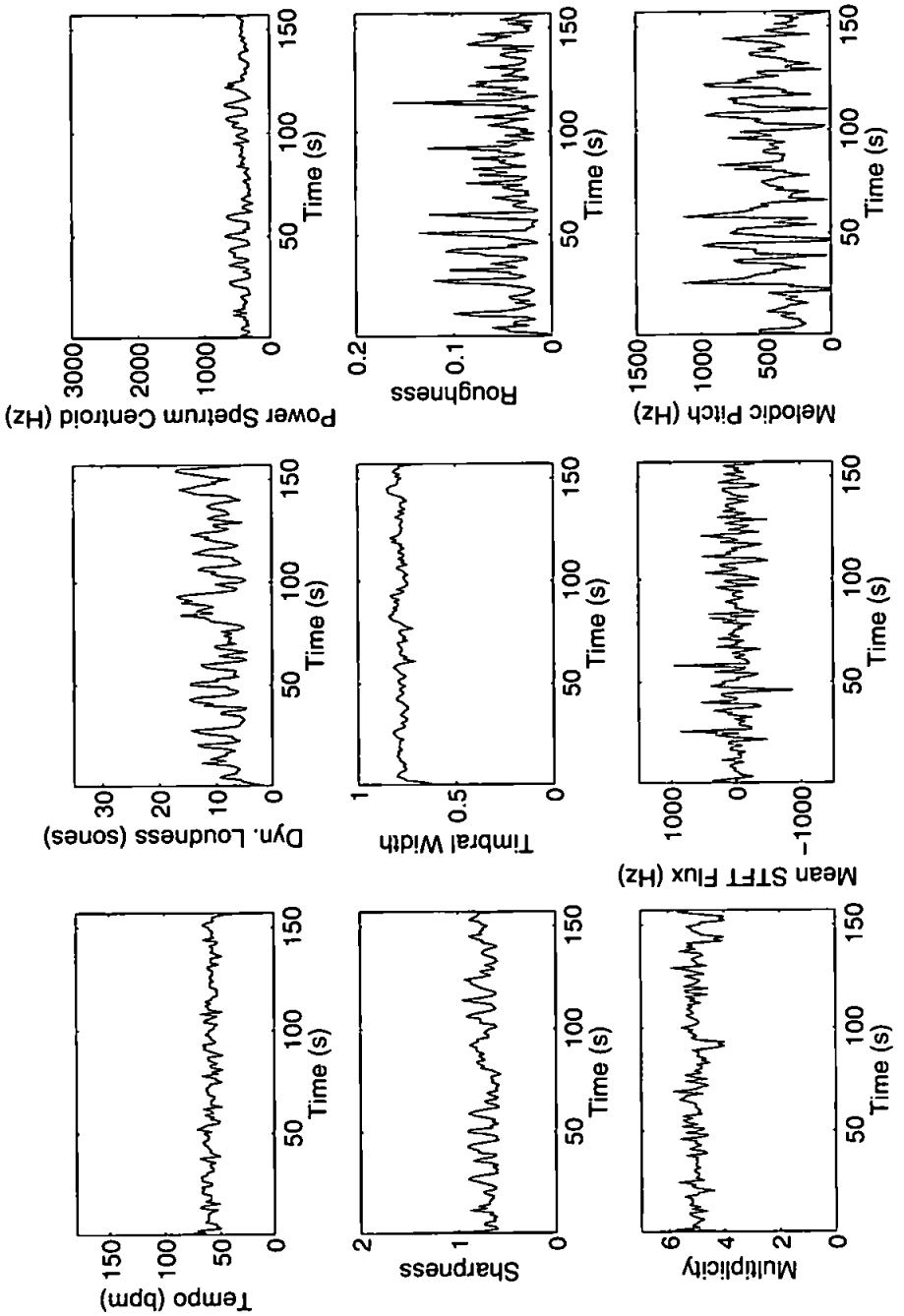


Figure F.6: Sound features extracted with Psysound 3: W. A. Mozart - Divertimento - II. "Allegro di molto" (piece 6)

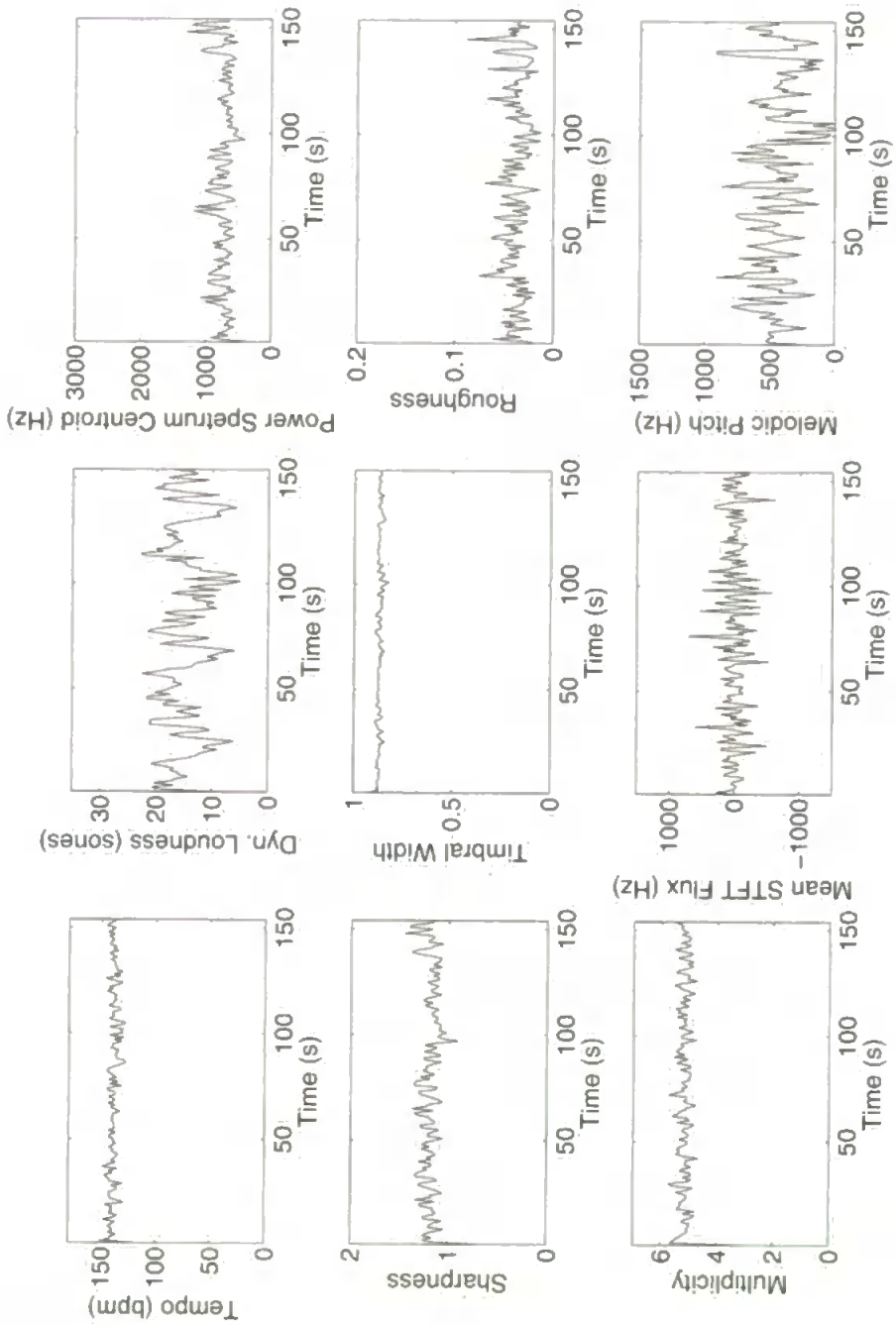


Figure F.7: Sound features extracted with Psysound 3: C. Debussy - La Mer - II. "Jeux de vagues" (piece 7)

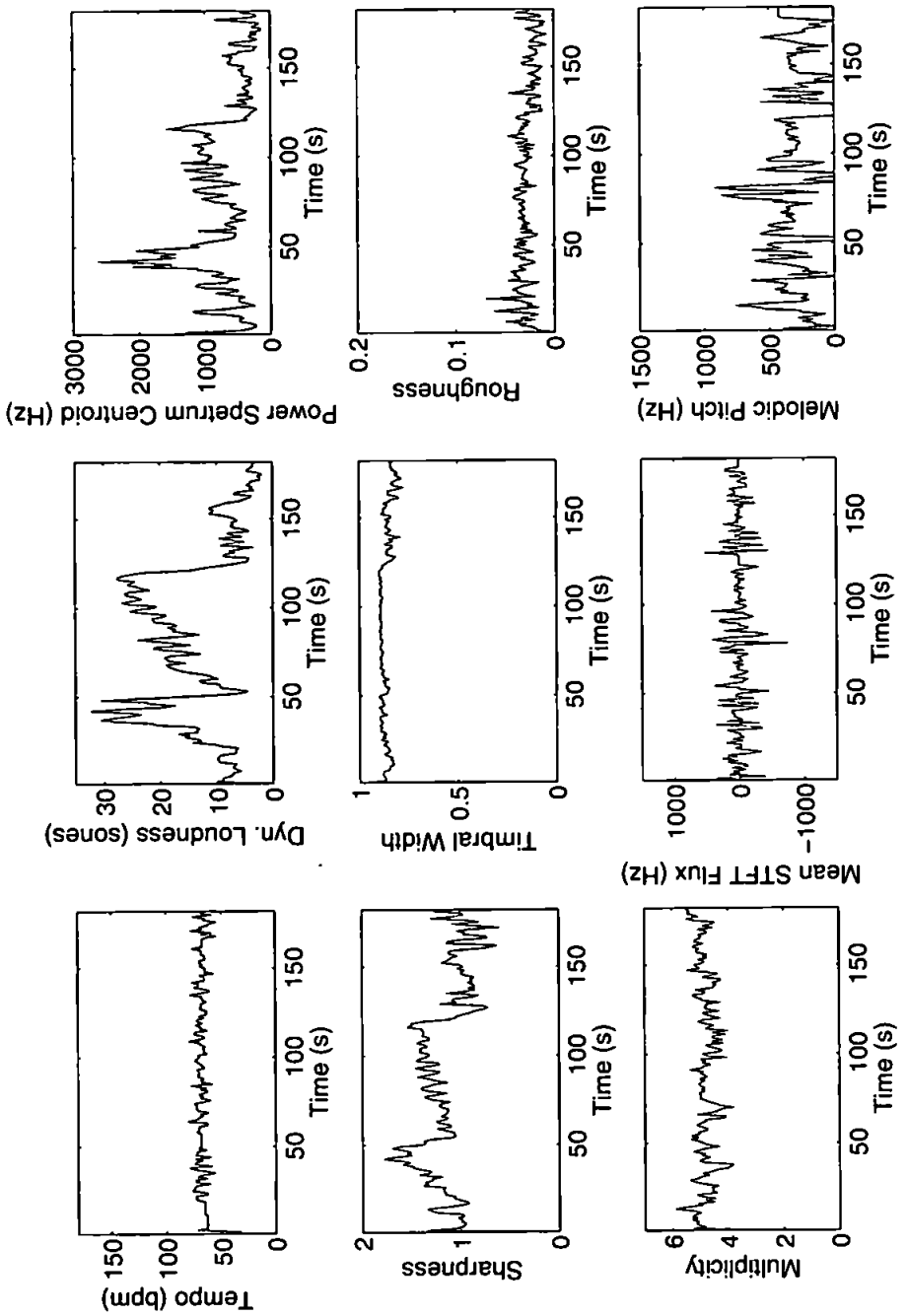


Figure F.8: Sound features extracted with Psysound 3: F. Liszt - Liebestraum No.3 (piece 8)

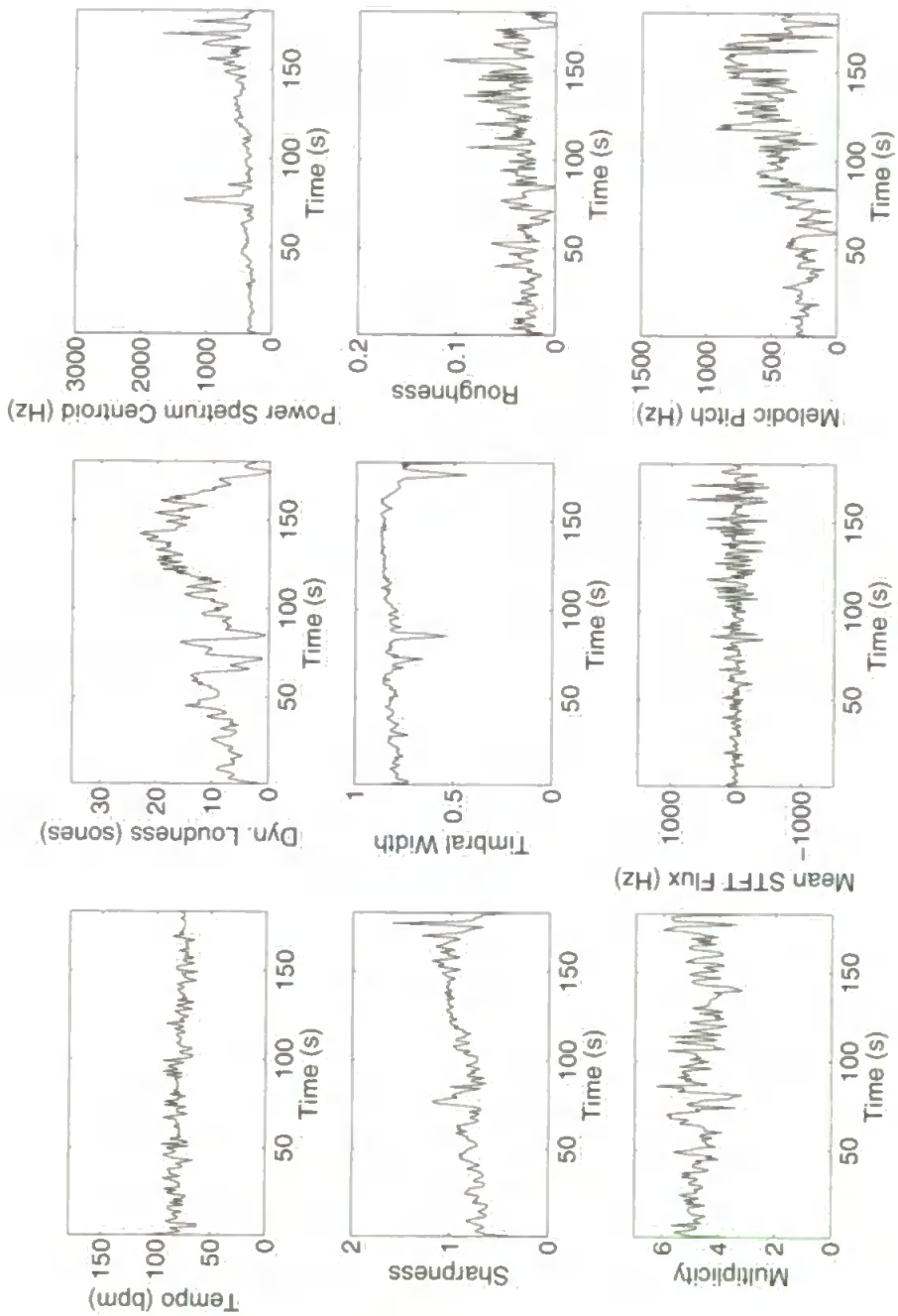
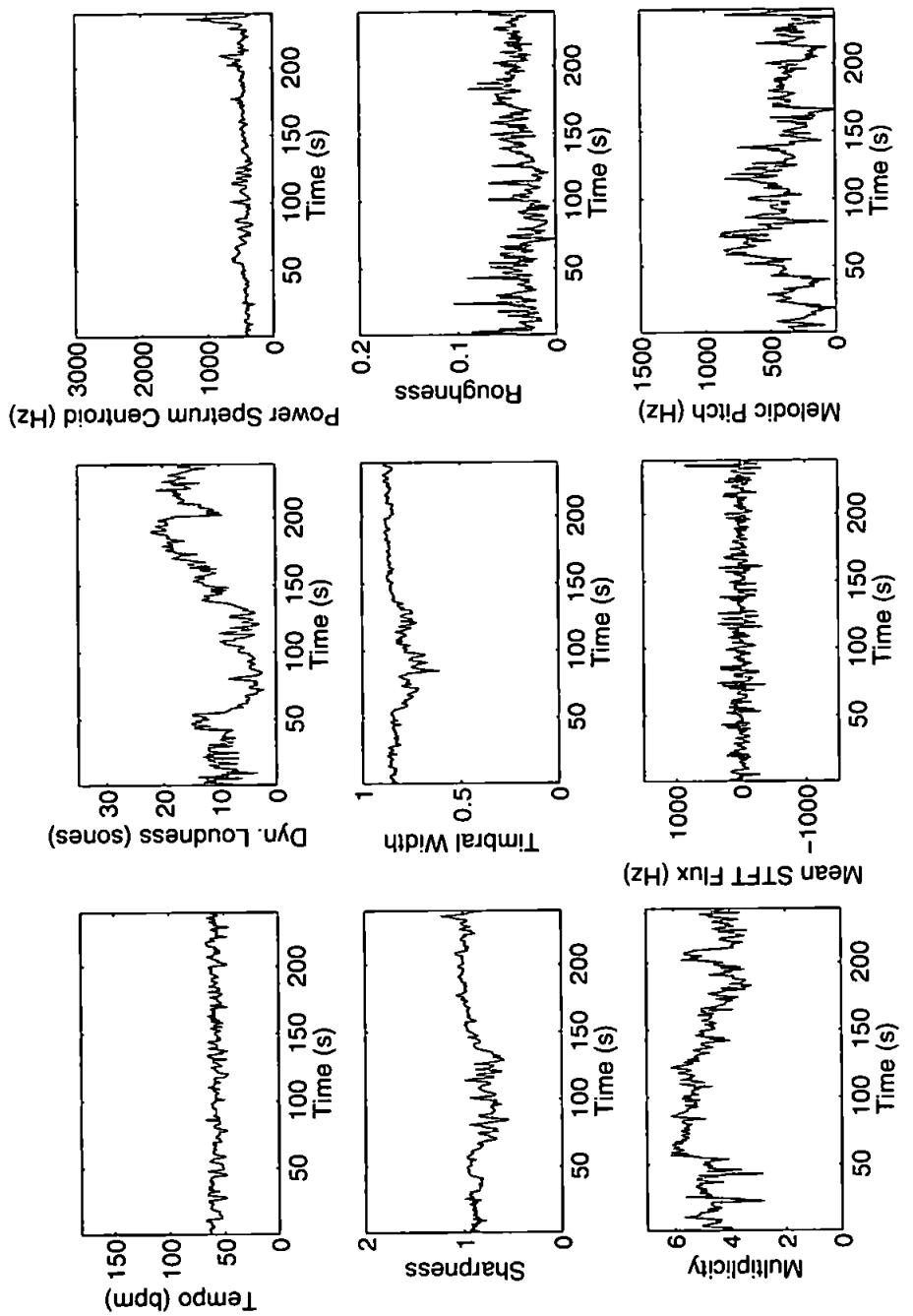


Figure F.9: Sound features extracted with Psysound 3: J. S. Bach - Partita No. 2 - "Chaconne" (piece 9)



List of references.

Arnold, M. (1960). *Emotion and Personality.* New York (NY, USA): Columbia University Press.

Asmus, E. (1985). The Development of a Multidimensional Instrument for the Measurement of Affective Responses to Music. *Psychology of Music*, *13*(1), 19.

Bagri, A., Sandner, G., & Di Scala, G. (1991). Wild running and switch-off behavior elicited by electrical stimulation of the inferior colliculus: effect of anticonvulsant drugs. *Pharmacology Biochemistry and Behavior*, *39*(3), 683-688.

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*(1), 43-64.

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.

Ben-Shakhar, G., Gati, I., Ben-Bassat, N., & Sniper, G. (2000, Jan). Orienting response reinstatement and dishabituation: effects of substituting, adding, and deleting components of nonsignificant stimuli. *Psychophysiology*, *37*(1), 102–10.

Berlyne, D. (1974). *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation.* London (UK): Halsted Press.

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.

Bharucha, J. (2002). Neural Nets, Temporal Composites, and Tonality.

Foundations of Cognitive Psychology: Core Readings.

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

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

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.

Bresin, R. (1998). Artificial neural networks based models for automatic performance of musical scores. *Journal of New Music Research*, *27*(3), 239–270.

Byers, P. (1976). Biological rhythms as information channels in interpersonal communication behavior. *Perspectives in Ethology*, *2*, 135–164.

Cabrera, D. (1999). Psysound: A computer program for psychoacoustical analysis. *Proceedings of the Australian Acoustical Society Conference*, *24*, 47-54.

Cabrera, D. (2000). *Psysound2: Psychoacoustical software for macintosh ppc.* July.

Cacioppo, J., Berntson, G., Larsen, J., Poehlmann, K., & Ito, T. (1993). The psychophysiology of emotion. *Handbook of emotions*, *2*, 119–142.

Cambouropoulos, E. (2003). Pitch Spelling: A Computational Model. *Music Perception*, *20*(4), 411–429.

Campbell, I. (1942). Basal emotional patterns expressible in music. *American Journal of Psychology*, *55*, 1–17.

Cannon, W. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *American Journal of Psychology*, *39*(1927), 106–

124.

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

Clynes, M. (1978). *Sentics: the touch of emotions.* New York: Doubleday.

Cohen, A. (1990). Understanding musical soundtracks. *Empirical Studies of the Arts*, *8*(2).

Cope, D. (1991). *Computers and musical style.* Oxford, UK: Oxford University Press.

Coutinho, E., Gimenes, M., Martins, J., & Miranda, E. (2005). Computational musicology: An artificial life approach. In *Proceedings of the 2nd portuguese workshop on artificial life and evolutionary algorithms workshop.* Covilha (Portugal): Springer Verlag.

Cuddy, L., & Lunney, C. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics*, *57*(4), 451-462.

Dahlstedt, P., & Nordhal, M. G. (2001). Living melodies: Coevolution of sonic communication. *Leonardo*, *34*(2), 243-248.

Dainow, E. (1977). Physical effects and motor responses to music. *Journal of Research in Music Education*, *25*(3), 211-221.

Damasio, A. (1994). *Descarte's error: emotion, reason and the human brain.* New York: Grosset/Putnam Books.

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

Davidson, R., Scherer, K., & Goldsmith, H. (2003). *Handbook of affective sciences.* USA: Oxford University Press.

Degazio, B. (1999). La evolucion de los organismos musicales. In E. R. Miranda (Ed.), *Musica y nuevas tecnologias: Perspectivas para el siglo xxi* (p. 137-148). Barcelona, Spain: L'Angelot.

Dell, G. (2002). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psycholinguistics: Critical Concepts in Psychology*, *93*(3), 283–321.

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

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* (p. 58-65). Ilmenau (Germany).

Dixon, S. (2001). Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*.

Dodge, C., & Jerse, T. (1985). *Computer music*. London, UK: Schimer Books.

Dowling, W., & Harwood, D. (1986). *Music cognition*. San Diego (CA), USA: Academic Press.

Downey, J. (1897). A musical experiment. *American Journal of Psychology*, *9*, 63-69.

Ekman, P. (1973). Cross-cultural studies of facial expression. In *Darwin and facial expression. a century of research*. New York (NY, USA): Academic Press.

Ekman, P. (1999). Basic emotions. In *The handbook of cognition and emotion* (p. 45-60). Sussex, UK: Wiley.

Ekman, P., & Davidson, R. (1994). *The Nature of Emotion: Fundamental Questions*. Cambridge, USA: Oxford University Press.

Ekman, P., Levenson, R., & 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.

Elman, J. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*(2-3), 195–225.

Farnsworth, P. (1958). *The social psychology of music*. New York (NY, USA): Dryden Press.

Feldman, L. (1995). Valence focus and arousal focus: individual differences in the structure of affective experience. *Journal of personality and social psychology*, *69*(1), 153-166.

Feng, Y., Zhuang, Y., & Pan, Y. (2003). Music information retrieval by detecting mood via computational media aesthetics. *Proceedings of the IEEE/WIC International Conference on Web Intelligence (WI 2003)*, 235–241.

Flowers, P. (1983). The Effect of Instruction in Vocabulary and Listening on Nonmusicians. *Journal of Research in Music Education*, *31*(3), 179–89.

Flowers, P. (1988). The Effects of Teaching and Learning Experiences, Tempo, and Mode on Undergraduates. *Journal of Research in Music Education*, *36*(1), 19–34.

Frijda, N. (1986). *The Emotions*. London, UK: Cambridge University Press.

Gabrielsson, A. (1991). Experiencing music. *Canadian Journal of Research in Music Education*, *33*, 21-26.

Gabrielsson, A. (2002). Emotion perceived and emotion felt: Same or different. *Musicae Scientiae*, 2001–2002.

Gabrielsson, A., & Juslin, P. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*(1), 68.

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In P. Juslin & J. Sloboda (Eds.), *Music and emotion: theory and research* (p. 223-248). New York: Oxford University Press.

Gabrielsson, A., & Lindström, S. (1995). Can strong experiences of music have therapeutic implications? In I. Steinberg (Ed.), *Music and the mind machine: The psychophysiology and psychopathology of the sense of music* (p. 195-202). Berlin: Springer Verlag.

Gaver, W., & Mandler, G. (1987). Play it again, Sam: On Liking Music. *Cognition & Emotion*, *1*(3), 259–282.

Giles, C., Lawrence, S., & Tsoi, A. (2001, July/August). Noisy time series prediction using a recurrent neural network and grammatical inference. *Machine Learning*, *44*(1/2), 161-183.

Gilman, B. (1891). Report of an experimental test of musical expressiveness. *Journal of Psychology*, *4*, 558-576.

Gilman, B. (1892). Report of an experimental test of musical expressiveness (continued). *Journal of Psychology*, *5*, 42-73.

Gimenes, M., & Miranda, E. (2008). An a-life approach to machine learning of musical worldviews for improvisation systems. In *Proceedings of 5th sound and music computing conference* (pp. 235–241). Berlin, Germany.

Giomo, C. (1993). An Experimental Study of Children's Sensitivity to Mood in Music. *Psychology of Music*, *21*(2), 141.

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

Goto, M. (2004). A real-time music-scene-description system: predominant-F0 estimation for detecting melody and bass lines in real-world audio signals. *Speech Communication*, *43*(4), 311–329.

Gray, P., & Wheeler, G. (1967). The semantic differential as an instrument to examine the recent folksong movement. *J Soc Psychol*, *72*(2), 241-247.

Greenwald, M., Cook, E., & Lang, P. (1989). Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of Psychophysiology*, *3*(1), 51–64.

Grewe, O., Nagel, F., Kopiez, R., & Altenmuller, 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.

Grewe, O., Nagel, F., Kopiez, R., & Altenmuller, E. (2007). Listening To Music As A Re-Creative Process: Physiological, Psychological, And

Psychoacoustical Correlates Of Chills And Strong Emotions. *Music Perception*, *24*(3), 297–314.

Guhn, M., Hamm, A., & Zentner, M. (2007). Physiological and Musico-Acoustic Correlates of the Chill Response. *Music Perception*, *24*(5), 473–484.

Gundlach, R. (1935). Factors determining the characterization of musical phrases. *American Journal of Psychology*, *47*, 624–644.

Hampton, P. (1945). The emotional element in music. *Journal of General Psychology*, *33*, 237–250.

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. london: William heinemann medical books* (p. 202-216). London, UK: Heinemann Medical.

Heinlein, C. (1928). The affective characters of the major and minor modes in music. *Journal of Comparative Psychology*(8), 101-142.

Hevner, K. (1936, April). 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.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780.

Hopfield, J. (1982). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences*, *79*(8), 2554–2558.

Hotelling, H. (1936). Relations between two sets of variables. *Biometrika*, *28*(3), 321–377.

Hunt, A., Kirk, R., & Orton, R. (1991). Musical applications of a cellular automata workstation. In *International computer music conference icmc91* (p. 165-168). San Francisco (CA), USA: ICMA.

Huron, D. (2001). Is music an evolutionary adaptation? *Annals of the New York Academy of Sciences*, *930*(1), 43-61.

Huron, D. (2006). *Sweet anticipation*. Cambridge (MA), USA: MIT Press.

Impett, J. (2001). Interaction, simulation and invention: a model for interactive music. In E. Bilotta, E. Miranda, P. Pantano, & P. Todd (Eds.), *Proceedings of almma 2001 workshop on artificial models for musical applications* (p. 108-119). Cosenza, Italy: Editoriale Bios.

Iwanaga, M., & Tsukamoto, M. (1997). Effects of excitative and sedative music on subjective and physiological relaxation. *Perceptual and Motor Skills*, *85*(1), 287-96.

Izard, C. (1979). *Emotions in personality and psychopathology*. Plenum Press.

Izard, C., Huebner, R., Risser, D., McGinnes, G., & Dougherty, L. (1980). The young infant's ability to produce discrete emotion expressions. *Developmental Psychology*, *16*(2), 132–140.

James, W. (1884). What is an emotion? *Mind*, *9*, 188-205. Available from `http://psychclassics.yorku.ca/James/emotion.htm` `(Last visited: 08 December 2008)`

Janata, P., & Grafton, S. (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nature Neuroscience*, *6*(7), 682–687.

Jeong, J., Joung, M., & Kim, S. (1998). Quantification of emotion by nonlinear analysis of the chaotic dynamics of electroencephalograms during perception of 1/f music. *Biological Cybernetics*, *78*(3), 217–225.

Jordan, M. (1990). Attractor dynamics and parallelism in a connectionist sequential machine. , 112–127.

Juslin, P., Friberg, A., & Bresin, R. (2002). Toward a computational model of expression in music performance: The GERM model. *Musicae Scientiae*, *6*(1), 63–122.

Juslin, P., & Laukka, P. (2003). Emotional Expression in Speech and Music Evidence of Cross-Modal Similarities. *Annals of the New York Academy of Sciences*, *1000*(1), 279–282.

Juslin, P., & Sloboda, J. (2001). *Music and emotion: theory and research.* New York, USA: Oxford University Press.

Kellaris, J., & Kent, R. (1993). An Exploratory Investigation of Responses Elicited by Music Varying in Tempo, Tonality, and Texture. *Journal of Consumer Psychology*, *2*(4), 381–401.

Khalfa, S., Peretz, I., Blondin, J., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, *328*(2), 145-149.

Kivy, P. (1989). *Sound Sentiment: An Essay on the Musical Emotions, Including the Complete Text of The Corded Shell.* Philadelphia, PA: Temple University Press.

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

Koelsch, S. (2005, Dec). Investigating emotion with music: neuroscientific approaches. *Ann N Y Acad Sci*, *1060*, 412-418.

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

Kohonen, T. (1977). *Associative Memory: A System-Theoretical Approach.* Berlin: Springer-Verlag.

Korhonen, M. (2004a). *Modeling continuous emotional appraisals of music using system identification.* Unpublished master's thesis, University of Waterloo.

Korhonen, M. (2004b, August). *Modeling continuous emotional appraisals of music using system identification.* online. (http://www.sauna.org/kiulu/emotion.html (Last visited 8 December 2008))

Korhonen, M., Clausi, D., & Jernigan, M. (2004, August). Modeling

continuous emotional appraisals using system identification. In R. O. G. P. W. S. D. Libscomb R. Ashley (Ed.), *Proceedings of the 8th international conference on music perception and cognition.* Evanston, USA.

Kratus, J. (1993). A Developmental Study of Children's Interpretation of Emotion in Music. *Psychology of Music*, *21*(1), 3.

Kremer, S. (2001). Spatiotemporal connectionist networks: A taxonomy and review. *Neural Computation*, *13*(2), 249-306.

Krumhansl, C. L. (1991). Melodic structure: Theoretical and empirical descriptions. In J. Sundberg (Ed.), *Music, language, speech and brain.* London: MacMillian.

Krumhansl, C. L. (1996). A perceptual analysis of mozart's piano sonata k. 282: Segmentation, tension, and musical ideas. *Music Perception*, *13*(3), 401–432.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*, *51*(4), 336-353.

Lang, P. (1979). A Bio-Informational Theory of Emotional Imagery. *Psychophysiology*, *16*(6), 495–512.

Lang, P., Bradley, M., & Cuthbert, B. (1998, September). Emotion and motivation: measuring affective perception. *Journal of Clinical Neurophysiology*, *15*(5), 397-408.

Lang, P., Bradley, M., & Cuthbert, B. (2005). *International affective picture system (iaps): Affective ratings of pictures and instruction manual* (Technical Report No. A-6). Gainesville, FL.: University of Florida.

Langer, S. (1942). *Philosophy in a new key.* Cambridge, USA: Harvard University Press.

Laukka, P. (2004). *Vocal expression of emotion: Discrete-emotions*

and dimensional accounts. Unpublished doctoral dissertation, Uppsala University, Sweden.

Lawrence, S., Giles, C., & Fong, S. (2000, Jan/Feb). Natural Language Grammatical Inference with Recurrent Neural Networks. *IEEE Transactions on Knowledge and Data Engineering*, *12*(1), 126-140.

Lazarus, R. (1991). Cognition and motivation in emotion. *American Psychologist*, *46*(4), 352–67.

Lee, V. (1932). *Music and its Lovers.* New York: E.P. Dutton.

Levitin, D. (2006). *This is your brain on music.* New York, USA: Dutton.

Li, T., & Ogihara, M. (2003). Detecting emotion in music. *Proceedings of the Fifth International Symposium on Music Information Retrieval*, 239-240.

Livingstone, S. (2007). Controlling musical emotionality: an affective computational architecture for influencing musical emotions. *Digital Creativity*, *18*(1), 43–53.

Ljung, L. (1986). *System identification: theory for the user.* New Jersey, USA: Prentice-Hall.

Lyon, R. (1984). Computational models of neural auditory processing. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, *9*.

Malloch, S. (1997). *Timbre and technology.* Unpublished doctoral dissertation, University of Edinburgh.

Mandelis, J. (2001). Genophone: An evolutionary approach to sound synthesis and performance. In E. Bilotta, E. Miranda, P. Pantano, & P. Todd (Eds.), *Proceedings of almma 2001 workshop on artificial models for musical applications* (p. 37-50). Cosenza, Italy: Editoriale Bios.

Mandelis, J., & Husbands, P. (2003). Musical interaction with artificial life forms: Sound synthesis and performance mappings. *Contemporary Music Review*, *22*(3), 69-77.

Manstead, A. (1988). The role of facial movement in emotion. *Social*

Psychophysiology: Theory and clinical applications, 105–129.

Manzolli, J., Moroni, A., Zuben, F. von, & Gudwin, R. (1999). An evolutionary approach applied to algorithmic composition. In E. R. Miranda & G. L. Ramalho (Eds.), *Vi brazilian symposium on computer music* (p. 201-210). Rio de Janeiro, Brazil: EntreLugar.

Mathews, M. (1963). The digital computer as a music instrument. *Science*, *142*(11), 553-557.

McAlpine, K., Miranda, E. R., & Hogar, S. (1999). Composing music with algorithms: A case study system. *Computer Music Journal*, *223*(2), 19-30.

McClelland, J., & Elman, J. (1986). Interactive processes in speech perception: the TRACE model. *Computational Models of Cognition And Perception*, 58–121.

McIntosh, D. (1996, June). Facial feedback hypotheses: Evidence, implications, and directions. *Motivation and Emotion*, *20*(2), 79-115.

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

McLaughlin, T. (1970). *Music and Communication.* London: Faber and Faber.

Meyer, L. (1956). *Emotion and meaning in music.* Imprint unknown.

Millen, D. (1990). Cellular automata music. In S. Arnold & D. Hair (Eds.), *International computer music conference icmc90* (p. 314-316). San Francisco (CA), USA: ICMA.

Miranda, E. (2004). At the crossroads of evolutionary computation and music: Self-programming synthesizers, swarm orchestras and the origins of melody. *Evolutionary Computation*, *12*(2), 137-158.

Miranda, E., Kirby, S., & Todd, P. (2003). On computational models of the evolution of music: From the origins of musical taste to the emergence of grammars. *Contemporary Music Review*, *22*(3), 91-111.

Miranda, E. R. (1993). Cellular automata music: An interdisciplinary music

project. *Interface (Journal of New Music Research)*, *22*(1), 03-21.

Mozer, M. (1999). Neural network music composition by prediction: Exploring the benefits of psychoacoustic constraints and multiscale processing. *Musical Networks: Parallel Distributed Perception and Performance.*

Mull, H. (1949). A study of humor in music. *Am J Psychol*, *62*(4), 560–6.

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.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication-realization model.* University Of Chicago Press.

Nelson, D. (1985). Trends in the Aesthetic Responses of Children to the Musical Experience. *Journal of Research in Music Education*, *33*(3), 193–203.

Nielsen, F. (1987). Musical tension and related concepts. In T. A. S. . J. Umiker-Seboek (Ed.), *The semiotic web* (p. 491-513). Berlin: Mouton de Gruyter.

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

Oliveira, A., & Cardoso, A. (2008, July). Controling music affective content: A symbolic approach. In C. Tsougras & R. Parncutt (Eds.), *Proceedings of the fourth conference on interdisciplinary musicology (cim08).* Thessaloniki, Greece.

Panksepp, J. (1995). The emotional sources of "chills" induced by music. *Music Perception*, *13*(2), 171-207.

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

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

Panksepp, J., & Bishop, P. (1981). An autoradiographic map of (3h) diprenorphine binding in rat brain: effects of social interaction. *Brain Research Bulletin*,

7(4), 405.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach.* Berlin: Springer Verlag. (Page 92)

Patel, A., & Balaban, E. (2000). Temporal patterns of human cortical activity reflect tone sequence structure. *Nature*, *404*, 80-84.

Peretz, I. (2001). Brain Specialization for Music: New Evidence from Congenital Amusia. *Annals of the New York Academy of Sciences*, *930*(1), 153.

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

Philippot, P., Chapelle, G., & Blairy, S. (2002). Respiratory feedback in the generation of emotion. *Cognition & Emotion*, *16*(5), 605-627.

Psysound 3. (2008). online. Available from `http://people.arch.usyd.edu.au/~sfer9710/PsySound3/` (Last visited: 08 December 2008)

Reisenzein, R. (1983). The Schachter theory of emotion: Two decades later. *Psychological Bulletin*, *94*(2), 239–264.

Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychology of Music.*

Rigg, M. (1937). Musical expression: An investigation of the theories of Erich Sorantin. *Journal of Experimental Psychology*, *21*, 442–455.

Rigg, M. (1964). The mood effects of music: a comparison of data from four investigators. *Journal of Psychology*, *58*, 427–438.

Rockstroh, B., Johnen, M., Elbert, T., Lutzenberger, W., Birbaumer, N., Rudolph, K., et al. (1987, Dec). The pattern and habituation of the orienting response in man and rats. *Int J Neurosci*, *37*(3-4), 169–82.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–

408.

Rosenthal, D., & Okuno, H. (1998). *Computational auditory scene analysis.* Mahwah, NJ: Lawrence Erlbaum Associates.

Rumelhart, D., Hintont, G., & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533-536.

Rumelhart, D., & McClelland, J. (1986). PDP models and general issues in cognitive science. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition (vol. 1).* Cambridge, MA, USA: MIT Press.

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

Russell, J. (1989). Measures of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience* (Vol. 4). Toronto, Canada: Academic.

Schachter, S. (1964). The interaction of cognitive and physiological determinants of emotional state. In N. Y. A. Press. (Ed.), *Advances in experimental social psychology* (p. 49-79). New York: Berkowitz.

Scherer, K. (1984). On the nature and function of emotion: a component process approach. In K. Scherer & P. Ekman (Eds.), *Approaches to emotion* (p. 293-317). Hillsdale, NJ (USA): Erlbaum.

Scherer, K. (1999). On the Sequential Nature of Appraisal Processes: Indirect Evidence from a Recognition Task. *Cognition & Emotion*, *13*(6), 763–793.

Scherer, K. (2004). Which emotions can be induced by music? what are the underlying mechanisms? and how can we measure them? *Journal of New Music Research*, *33*(3), 239-251.

Scherer, K., & Oshinsky, J. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*(4), 331-346.

Scherer, K., & Zentner, M. (2001). Emotional effects of music: Production rules.

Music and emotion: Theory and research, 361-392.

Schubert, E. (1999a). *Measurement and time series analysis of emotion in music.* Unpublished doctoral dissertation, Univ. of New South Wales.

Schubert, E. (1999b). Measuring emotion continuously: Validity and reliability of the two dimensional emotion space. *Australian Journal of Psychology*, *51*(3), 154-165.

Schubert, E. (2004). Modeling Perceived Emotion With Continuous Musical Features. *Music Perception*, *21*(4), 561–585.

Schubert, E., & Dunsmuir, W. (1999). Regression modelling continuous data in music psychology. In S. W. Yi (Ed.), *Music, mind, and science* (p. 298-352). Seoul National University Press.

Sherman, M. (1928). Emotional Character in the Singing Voice. *Journal of Experimental Psychology*, *11*, 495–497.

Sloboda, J. (1991). Music structure and emotional response: some empirical findings. *Psychology of Music*, *19*, 110-120.

Sloboda, J., & Lehmann, A. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a chopin piano prelude. *Music Perception*, *19*(1), 87-120.

Stuart, L., Marocco, D., & Cangelosi, A. (2005). Information visualization for knowledge extraction in neural networks. In W. Duch, J. Kacprzyk, E. Oja, & S. Zadrozny (Eds.), *Artificial neural networks: Formal models and their applications (proceedings of the 15th icann 2005)* (Vol. 3697/2005, p. 515-520). Berlin (Germany): Springer.

Suzuki, K., & Hashimoto, S. (1997). Modeling of emotional sound space using neural networks. *Proc. Intl Workshop on KANSEI: The technology of emotion, AIMI and DIST-University of Genova, Genova*, 116-121.

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

of Arts and Science.

Tillmann, B., Bharucha, J., & Bigand, E. (2000). Implicit Learning of Tonality: A Self-Organizing Approach. *Psychological Review*, *107*(4), 885–913.

Todd, N. (1992). The dynamics of dynamics: A model of musical expression. *The Journal of the Acoustical Society of America*, *91*(6), 3540–3550.

Todd, P., & Werner, G. (1999). Frankensteinian methods for evolutionary music composition. In N. Griffith & P. Todd (Eds.), *Musical networks: Parallel distributed perception and performance* (p. 313-339). Cambridge (MA), USA: MIT Press/Bradford Books.

Tomkins, S. (1980). Affect as amplification: Some modifications in theory. In R. P. . H. Kellerman (Ed.), *Theories of emotion* (pp. 141–164). New York: Academic Press.

Tzanetakis, G., & Cook, P. (2000). Marsyas: a framework for audio analysis. *Organised Sound*, *4*(03), 169-175.

Valentine, C. (1914). The method of comparison in experiments with musical intervals and the effects of practice on the appreciation of discords. *British Journal of Psychology*, *7*, 118–135.

Valentine, C. (1962). *The experimental psychology of beauty*. London: Methuen.

Vanderark, S., & Ely, D. (1993). Cortisol,biochemical, and galvanic skin responses to musical stimuli of different preference values by college students in biology and music. *Perceptual and Motor Skills*, *77*, 227-234.

Voss, R., & Clarke, J. (1978). –1/f noise–in music: Music from 1/f noise. *The Journal of the Acoustical Society of America*, *63*, 258.

Waschka II, R. (1999). Avoiding the fitness bottleneck: Using genetic algorithms to compose orchestral music. In *International computer music conference icmc99* (p. 201-203). San Francisco (CA), USA: ICMA.

Washburn, M., & Dickinson, G. (1927). The sources and nature of the affective reaction to instrumental music. In M. Schoen (Ed.), *The effects of music: a*

series of essays (p. 121-130). New York, USA: Harcourt Brace.

Waterman, M. (1996). Emotional Responses to Music: Implicit and Explicit Effects in Listeners and Performers. *Psychology of Music*, *24*(1), 53.

Watson, K. (1942). The nature and measurement of musical meanings. In *Psychological monographs* (Vol. 244). The American psychological association.

Wedin, L. (1969). Dimension analysis of emotional expression in music. *Swedish Journal of Musicology*, *16*, 118–140.

Weld, H. (1912). An experimental study of musical enjoyment. *American Journal of Psychology*, *23*, 245–308.

White, P. (2000). *Basic midi.* London, UK: Sanctuary Publishing.

Widmer, G., & Goebl, W. (2004). Computational Models of Expressive Music Performance: The State of the Art. *Journal of New Music Research*, *33*(3), 203–216.

Widrow, B., & Hoff, M. (1960). Adaptive switching circuits. 1960 IRE WESCON Conv. *Record, Part*, *4*, 96–104.

Witvliet, C. v. O., & Vrana, S. R. (1996). The emotional impact of instrumental music on affect ratings, facial emg, autonomic measures, and the startle reflex: effects of valence and arousal. *Psychophysiology Supplement*, *91*.

Witvliet, C. v. O., Vrana, S. R., & Webb-Talmadge, N. (1998). In the mood: Emotion and facial expressions during and after instrumental music and during an emotional inhibition task. *Psychophysiology Supplement*, *88*.

Worral, D. (2001). Studies in metamusical methods for sound image and composition. *Organised Sound*, *1*(3), 183-194.

Wundt, W. (1896). *Grundriss der psychologie (outlines of psychology).* Leipzig: Engelmann.

Xenakis, I. (1971). *Formalized music: Thought and mathematics in composition.* Bloomington (IN), USA: Indiana University Press.

Zajonc, R., & McIntosh, D. (1992, Jan). Emotions research: Some promising questions and some questionable promises. *Psychological Science.*

Zatorre, R. J. (2005). Music, the food of neuroscience? *Nature, 434*, 312-315.

Zentner, M., Meylan, S., & Scherer, K. (2000). Exploring musical emotions across five genres of music. *6th International Conference of Society for Music Perception and Cognition, Keele, UK.*

Zillmann, D. (1983). Social psychophysiology. In C. . R. Petty (Ed.), *Social psychophysiology: A sourcebook* (pp. 215–240). New York: Guilford.

Zwicker, E., & Fastl, H. (1990). *Psychoacoustics.* Springer New York.

Bound copies of published papers.

Coutinho, E. & Cangelosi, A. (in press). The use of spatiotemporal connectionist models in psychological studies of musical emotions. *Music Perception.*

Running head: NEURAL NETWORK MODELS OF MUSICAL EMOTIONS

The use of spatio-temporal connectionist models in psychological studies of musical emotions

Eduardo Coutinho and Angelo Cangelosi

University of Plymouth

Plymouth, Devon

United Kingdom

Abstract

This article presents a novel methodology to analyze the dynamics of emotional responses to music. It consists of a computational investigation based on spatiotemporal neural networks, which are of "mimic" human affective responses to music and to predict the responses to novel music sequences. The results provide evidence suggesting that spatiotemporal patterns of sound resonate with affective features underlying judgments of subjective feelings (arousal and valence). A significant part of the listener's affective response is predicted from the a set of six psychoacoustic features of sound – loudness, tempo, texture, mean pitch, pitch variation and sharpness. A detailed analysis of the network parameters and dynamics also allows us to identify the role of specific psychoacoustic variables (e.g., tempo and loudness) in music emotional appraisal. This work contributes with new evidence and insights to the study of musical emotions, with particular relevance to the music perception and cognition research community.

Keywords: Emotion, Music, Arousal and Valence, Psychoacoustics, Neural Networks.

The use of spatio-temporal connectionist models in psychological studies of musical emotions

Ever since antiquity, the relationship between music and emotion has been acknowledged as a fascinating quality of the human experience. Ancient philosophers, such as Socrates, Plato, or Aristotle, in their theories of emotion, considered the sound of music and the unique way in which it can reflect "states of the soul". For the Greek philosophers music has the power to alter and drive the collective consciousness of massive groups of people.

Many years have passed and we still haven't found an answer to expose the mechanisms that music uses to interact with emotional systems. Nevertheless, the revival of studies on emotions during the late 19th century together with the new technological developments in measurement techniques, have contributed with new insights for such an old question: how does music affect emotions?

"Cognitivist" and "Emotivist" views

There are two principle, complementary views regarding the relationships between music and emotions. "Cognitivists" defend that music simply expresses emotions that the listener can identify, while "emotivists" defend that music can elicit affective responses in the listener (see Krumhansl, 1997; Kivy, 1990).

One of the most influential works from a "cognitivist" perspective was by Meyer (1956). He developed a theory in which musical emotions depend mainly upon expectations about the unfolding events and their meanings, which create patterns of tension and release in the listener (Meyer, 1956). For Meyer, expectation is a necessary condition for emotion and meaning to be conveyed in music. The nature of these expectations derives from the development of psychological schemas of systems of sound relationships. These include the general Gestalt principles for perceptual organization, but mainly psychological schemas derived from the

interaction with a given (musical) culture. Without the "stylistic experience" music becomes meaningless and consequently lacks in affect. Empirical support for Meyer's ideas has come from different formalizations of his theory (e.g., Narmour, 1992; Krumhansl, 1991; Cuddy & Lunney, 1995). Meyer's "cognivist" perspective is especially evident in a passage of his 1956 book: "... when a listener reports that he felt this or that emotion, he is describing the emotion which he believes the passage is supposed to indicate, not anything which he himself has experienced" (Meyer, 1956, p. 8). Although our affective experiences with music are ultimately individual and culturally dependent, "Emotivists" claim that music can itself elicit emotions in listeners. From this perspective, there are certain music dimensions and qualities which induce similar affective experiences in all listeners, cross-culturally, and independent of context and personal biases or preferences. Some evidence about the universality of music affect comes from a cross-cultural study by Balkwill and Thompson (1999). Western listeners (who had no familiarity with North Indian ragas) listened to Hindustani music and were able to identify emotions of joy, sadness, and peace.

More compelling evidence suggesting that music itself can elicit emotions without the involvement of cognition, favoring the "emotivist" view on musical emotions, can be found in Peretz, Gagnon, and Bouchard (1998). Peretz et al. (1998) described a patient (I.R.) suffering from severe loss of music recognition and expressive abilities. I.R. showed no evidence of impairment in the auditory system but still she couldn't discriminate pitch and temporal deviations in the music. Even violations of the scale structure, or judgments of adequacy of a pitch as the ending of a harmonic sequence (tonal closure) were impossible to I.R. Despite all this, I.R. still claims the capacity to enjoy music. In the experiment, the patient was able to derive the emotional tone of the excerpts, manipulated in terms of tempo and mode to achieve the

intended emotional qualities. Although I.R. was not aware of the music manipulations[1], she performed as well as the control group on the affective content identification task. This study shows that the perceptual analysis of the music input can be maintained for emotional purposes, even if impaired for cognitive ones. Peretz et al. (1998) suggest the possibility that emotional and nonemotional judgments are the products of distinct neurological pathways. Some of these pathways were found to involve the activation of sub-cortical 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 even outside an individuals' awareness. Panksepp and Bernatzky (2002) even suggest that a great part of the emotional power derived from music may be generated by lower subcortical regions, where basic affective states are organized (Damasio, 2000; Panksepp, 1998).

Taking together these findings provide evidence of the universality of music affect and that cognitive mediation is not a required element in music appreciation. But in that case, for the affective experience to happen, it is plausible to think that the listener must derive affective meaning from the nature of the stimulus. This approach follows the view advocated by Langer (1942) on the existence of expressive forms ("iconic symbols") of emotions in all art forms. She believed that the arts and music in particular, are fundamental forms of human physical and mental life.

Music elements and the construct of Emotion

One of the major obstacles for experimental studies on the emotional power of music is the subjective nature and multiple components of the affective experience. Nevertheless, by focusing on the time course of emotional responses to music, several experimental studies suggest some generalizations. One of the most important is that different listeners report

emotional responses to music, consistent in their quality and intensity. This led some studies to focus on the music features (e.g., tempo, mode, dynamics, among others), attributing to the variables correspondences with particular affective experiences. Some pioneering studies that investigated the influence of music parameters on perceived emotion were published by Hevner (1936). Hevner attempted a systematic explanation of such relationships. Since then a core interest amongst music psychologists has been the isolation and measurement of the perceptible factors in music, which may be responsible for the resultant affective value (Gabrielsson & Lindström, 2001). The belief is that the way the sound elements are chosen and organized in time is linked with the listeners' affective experience.

Much of the research in this area has focused on general emotional characterizations of music (e.g., identification of basic emotion, lists of adjectives, or affective labels), by controlling parameters that can show some degree of stability throughout a piece (e.g., tempo, key, timbre, mode). In some studies, sets of specially designed stimuli have been used (e.g., probe tone test), while other studies were based on a systematic manipulation of real music samples (e.g. slow down tempo, changing instruments). More recently, following the claim that music features and structure are characterized by emotionally meaningful changes over time (e.g., Dowling & Harwood, 1986), new frameworks starting to use real music and continuous measurements of emotion emerged (e.g., Schubert, 2001).

Schubert (2001) proposed the use of continuous measurements of cognitive self-report of emotion; using a dimensional paradigm to represent emotions on a continuous scale. According to Wundt (1896), differences in the affective meaning among stimuli can succinctly be described by three pervasive dimensions (of human judgment): pleasure ("lust"), tension ("spannung"), and inhibition ("beruhigung"). This model has received empirical support from several studies,

which have shown that a large spectrum of continuous and symbolic stimuli can be represented using these dimensions (see Bradley & Lang, 1994). They can be represented in a three-dimensional space, with each dimension corresponding to a continuous bipolar rating scale: pleasantness-unpleasantness, rest-activation, and tension-relaxation. Other studies have provided also evidence that the use of only two dimensions is a good framework to represent affective responses to linguistic (Russell, 1980), pictorial (Bradley & Lang, 1994), and music stimuli (Thayer, 1986). These dimensions are labeled as Arousal and Valence. Arousal corresponds to a subjective state of feeling activated or deactivated. Valence stands for a subjective feeling of pleasantness or unpleasantness (hedonic value) (Russell, 1989).

The use of dimensional models is by itself a limitation on the representation and measurement of music emotions. Principally, the limitation is due to the wide variety of emotions conveyed by music and their limited representation by such a model. Another limitation arises due to the focus placed on a limited characterization of emotion: by asking participants to focus on their feelings, other components of emotion are not controlled for. Nevertheless the model shows important advantages compared with other methods used (generally classified as discrete emotions and eclectic approaches (Scherer, 2004)). First, they are suitable to be used with continuous measurement frameworks. In this way, they allow to analyze the time course of emotion in more detail than the other methods. Second, because they describe a continuous space not attached to a specific label, they also allow for the representation of a very wide range of emotional states, which is especially important in the context of music. By acknowledging its disadvantages, and by considering the important advantages offered by this method (particularly the simplicity in terms of psychological experiments and good reliability

(Scherer, 2004)), dimensional approaches to emotion representation have been consistently used in emotion research.

Models of continuous measurements of emotion in music

Following a continuous measurement framework, some studies have focused on analyzing temporal patterns in music and emotion. The music stimuli are encoded into time-varying patterns in the form of psychoacoustic features. These correspond to perceptually separable elements (or groups of elements), that when combined provide a description of the "perceptual object". Their division into separable sound dimensions allows for the study of their dynamics individually. The phenomena of music perception can then be described at different levels of detail, by selecting among different combinations of features.

Within this framework, two mathematical models that used time-varying patterns of music and emotion ratings have been proposed. Schubert (1999a) applied an ordinary least squares stepwise linear regression and a first order autoregressive model to his experimental data. He created regression models of emotional ratings, for selected music features, at different time lags for each piece. The relationships between music and emotional ratings were assumed to be linear and mutually independent, not accounting for the interactions among variables. The models also had the disadvantage of being piece specific.

Korhonen (2004) adopted a different modeling paradigm and extended the sound feature space and the music repertoire. He chose System Identification (Ljung, 1987) to model time-varying patterns of psychoacoustic features and emotion ratings. Korhonen's contributions are the integration of all music features into a single module and the possibility to use the model with unknown pieces. Despite some improvements over Schubert's work, the performance of this model is irregular. It outperformed Schubert's models for some pieces, but performed worse

in others. Another important disadvantage is that no insights on the processes used by the models to achieve the predictions are made. It is difficult to assess the meaningfulness of the relationships established to model the affective reactions based on sound features.

Although traditional time series analysis techniques allow for an investigation of the relationships between different processes, they often assume too much about the nature of the signals and their underlying behavior (due to assumptions like stationarity (Brockwell & Davis, 1991)). The work by Schubert (1999a) and Korhonen (2004) has shown the relevance of auto and cross correlations among psychoacoustic variables, and the limitations associated with the use of time series analysis techniques (e.g., pdf's, stationarity, linear correlations). In the two studies described, the model analysis only highlight positive relationships between tempo and loudness gradients and arousal ratings. Other observations derived from the model analysis often lack in generality.

Spatio-temporal connectionist networks

In order to overcome such limitations we suggest that spatio-temporal connectionist networks (Kremer, 2001) offer an ideal platform for the investigation of the dynamics of affective responses to music. Specifically we propose the use of recurrent neural networks. The fundamental additional aspect of this neural network (when compared with the traditional feed-forward model) is the use of recurrent connections that endow the network with a dynamic memory.

Various proposals and architectures can be found in literature for time-based neural networks (see Kremer, 2001 for a review), which make use of recurrent connections in different contexts. In our study we have selected the Elman network (Elman, 1990), also called Simple Recurrent Network. An Elman Neural Network (ENN) is based on the standard architecture of a

multi-layer perceptron with an additional "context" or "memory" layer. The units in this layer receive a copy of the previous internal state of the hidden layer. They are connected back to the same hidden layer, through adjustable weights. These units endow the network with a dynamic memory, achieved through recursive access to past information of internal representations of input stimuli.

The internal representations of an ENN encode not only the prior event but also relevant aspects of the representation that was constructed in predicting the prior event from its predecessor (that is the effect of having learned weights from the memory to the hidden layer). The basic functional assumption is that the next element in a time-series sequence can be predicted by accessing a compressed representation of previous hidden states of the network and the current inputs. 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. During learning, the hidden units must accomplish an input-output mapping and simultaneously develop representations that systematic encodings of the temporal properties of the sequential input at different levels (Elman, 1990). In this way, the internal representations that drive the outputs are sensitive to the temporal context of the task (even though the effect of time is implicit). The recursive nature of these representations (acting as an input at each time step) endows the network with the capability of detecting time relationships of sequences of features, or combinations of features, at different time lags (Elman, 1991). This is an important feature of this network because the lag between music and affective events has been consistently shown to vary over a range of five seconds (Schubert, 2004; Krumhansl, 1996; Sloboda & Lehmann, 2001).

ENNs use a training phase and a testing phase. Learning algorithms (supervised or unsupervised) define the way the model behaves during training when tuning its parameters for a certain task. The testing phase serves to test the model with novel data, either for prediction or validation of the model. A typical example of neural network training is categorization: given a set of training stimuli, the model is asked to separate them into a predetermined set of categories. An interesting phenomenon arises when we present the system with novel stimuli. These new inputs, after a successful learning process, should ideally be categorized within the learned categories space, reflecting the underlying grammar of the process being modeled. This process is called generalization and allows connectionist models to categorize novel stimuli.

In this article we will use an ENN to model continuous measurements of affective responses to music, based on a set of psychoacoustic components extracted from the music stimuli. Following the modeling stage, we make use of a set of analytical techniques, which allow for a better understanding of the relationships between sound features and affective responses. We then discuss the performance of our model and the implications of our findings for the "emotivist" and "cognitivist" perspectives on musical emotions.

Simulation Experiments

Method

The data for the experiments was obtained from a study conducted by Korhonen (2004)[2]. The original self-report data includes the emotional appraisals of six selections of classical music (see Table 1), obtained from 35 participants (21 male and 14 female). Using a continuous measurement framework, emotion was represented by its valence and arousal dimensions (using the EmotionSpace Lab (Schubert, 1999b)). The emotional appraisal data was collected at 1Hz (second-by-second).

– Insert Table 1 –

Encoded features

Korhonen encoded the music pieces into the psychoacoustic space by extracting low and high level features, using Marsyas (Tzanetakis & Cook, 1999) and PsySound (Cabrera D. , 2000) software packages. Only Tempo was calculated manually, using Schubert's method as explained in Schubert (1999a). The 13 psychoacoustic variables chosen (the 5 sound features representing Harmony variables included in Korhonen's study are not included here in order to exclude higher level features specific to the music culture and with controversial methods for its quantification) are shown in Table 2 and described below (for convenience we will refer to the input variables with the aliases indicated in this table). Because some of these measures refer to the same psychoacoustic dimension, they were clustered into 6 major groups: Dynamics, Mean Pitch, Pitch Variation, Timbre, Tempo, and Texture.

– Insert Table 2 –

Dynamics: The Loudness Level (D_1) and the Short Term Maximum Loudness (D_2) represent the subjective impression of the intensity of a sound (measured in sones). Both algorithms estimate the same quantity (described in Cabrera, 1999) and output similar values.

Mean Pitch: The Mean Pitch was quantified using two power spectrum calculations (one from PsySound, and another from Marsyas). The Power Spectrum Centroid (P_1) represents the first moment of the power spectral density (PSD) (Cabrera D. , 1999). The Mean STFT Centroid (P_2) is a similar measure and corresponds to the balancing point of the spectrum (Tzanetakis & Cook, 1999).

Pitch Variation: The pitch contour was quantified using 3 measures. The Mean STFT Flux (Pv_1) corresponds to the Euclidian norm of the difference between the magnitudes of the

Short Time Fourier Transform (STFT) spectrum evaluated at two successive sound frames. The standard deviation of P_2 (Pv_2) and of Pv_1 (Pv_3), were also used to quantify the pitch variations[3] (refer to Tzanetakis & Cook, 1999 for furthers details).

Timbre: Timbre was represented using the 4 different measures: Sharpness (Ti_1), a measure of the weighted centroids of the specific loudness, approximates the subjective experience of a sound on a scale from dull to sharp -the unit of sharpness is the acum (one acum is defined as the sharpness of a band of noise centered on 1000 Hz, 1 critical-bandwidth wide, with a sound pressure level of 60 dB) -details on the algorithm used in Psysound can be found in Zwicker & Fastl, 1990); Timbral Width (Ti_2) is a measure proposed by Malloch (1997) which measures the flatness of the specific loudness function, quantified as the width of the peak of the specific loudness spectrum (see Cabrera, 1999 for further details and slight modifications to that algorithm); the mean and standard deviations of the Spectral Roll-off (the point where a frequency that is below some percentage of the power spectrum resides -refer to Tzanetakis & Cook, 1999 for the detail on these measures) are also two measures of spectral shape (Ti_3 and Ti_4) -although they do not directly represent timbre, Korhonen included these measures because they have been successfully used in music information retrieval.

Tempo: Tempo was estimated from the number of beats per minute. Because the beats were detected manually a linear interpolation between beats was used to transform the data into second-by-second values (details on the tempo estimation are described in Schubert, 1999a).

Texture: Multiplicity (Tx) is an estimate of the number of tones simultaneously noticed in a sound; this feature was quantified using Parncutt's algorithm, as described in Parncutt (1989, p. 92) included in Psysound.

Modeling procedure

The psychoacoustic features constitute the input for our model. Each of these variables corresponds to a single input node of the network. The output layer consists of 2 nodes representing Arousal and Valence. Three pieces of music (1, 2, and 5), corresponding to 486s, were used during the training phase. In order to evaluate the response to novel stimuli, we used the remaining 3 pieces: 3, 4, and 6 (632s of music). Throughout this article we refer to the "Training set" as the collection of stimuli used to train the model, and "Test set" to the novel stimuli, unknown to the system during training, that test its generalization capabilities and performance. The task at each training iteration is to predict the next (t+1) values of Arousal and Valence. The target values (aka "teaching input") are the average Arousal/Valence pairs across all participants in Korhonen's experiments. In order to adapt the range of values of each variable to be used with the network, all variables were normalized to a range between 0 and 1.

The learning process was implemented using a standard back-propagation technique (Rumelhart, Hintont, & Williams, 1986). During training the same learning rate and momentum were used for each of the 3 connection matrices. The network weights were initialized with different random values. The range of values for each connection in the network (except for the connections from the hidden to the memory layer which are set constant to 1.0) was defined randomly between -0.05 and 0.05.

If the model is also 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. To avoid the over-fitting of the training set, we estimated the maximum number of training iterations and learning parameters. After preliminary tests and analysis, we decided upon 20000 iterations as the duration of training, using a learning rate of .075 and a momentum of 0. The size of the hidden layer (which defines the dimensionality of the internal

space of representations) was also optimized by testing the model with different numbers of hidden nodes. The best performance was obtained with a hidden layer of size five.

The root mean square (RMS) error is used here to quantify the differences between values predicted by the model and the values actually observed experimentally. Although this is a common measure to assess the performance of connectionist models, it gives little guaranties about a successful modeling process. We will use this measure only to compare the model performance with alternative sets of inputs to the network (next subsection). To assess the model ability to categorize the stimuli in terms of their affective value (and so the meaningfulness of the modeling process), we will analyze in detail the model categorization process.

Simulation 1: Reduction of the psychoacoustic (input) dimensions

The choice of the input space must consider musical, psychological and modeling aspects. The psychoacoustic features chosen by Korhonen include a significant set of perceptually relevant dimensions, although there are some redundancies to address. A recurrent problem in dealing with this type of data are the correlations among the encoded dimensions, especially redundant information and collinearity (as discussed by Schubert (1999a)). Because of that we decided to use only one variable of each of the psychoacoustic dimensions considered.

We started our simulations by training the neural network with different groups of inputs. Tempo, Texture, Dynamics, Mean Pitch, Pitch Variation, and Timbre are all considered to be included in the model as separate dimensions. In the case of Tempo and Texture, because they ere estimated using a single method (algorithm), they are included directly because there is no choice among alternative measures to be made[4]. In order to select one sound feature from the remaining music dimensions (Dynamics, Mean Pitch, Pitch Variation, and Timbre), each set of inputs considered included all unique features for each music dimension as a basic set (T and Tx

as explained before), plus one other test variable(s). For instance, in the case of Dynamics we tested T, Tx, D_1,and D_2[5], but also T, Tx, and D_1, and T, Tx, and D_2. We followed the same procedure for Mean Pitch, Pitch Variation, and Timbre.

For each test case we trained three different neural networks (with different random configuration of initial weights) and averaged their errors. In Table 3 are shown the RMS errors for each test condition.

– Insert Table 3 –

For the loudness measures, we found that the inclusion of both variables, or only D_1, produced the best results. We selected D_1 from this group. Regarding Timbre, the best performance was achieved using only Ti_1, and so this variable was also selected. The variable selected to represent Mean Pitch is P_1, because it performs better than the remaining variables. Finally, Pitch variation shows very similar error values for all test cases. We chose Pv_1 because it yields a lower error than Pv_2 and Pv_3.

We trained another network including all the variables chosen (T, Tx, D_1, P_1, Ti_1, and Pv_1) in order to assess the performance with all variables together. The results are shown at the bottom of Table 3. An inspection of the RMS error shows that combining all the features improved the model performance substantially, suggesting that the interaction among different features conveys relevant information. In the following simulation experiment, we will use the selected 6 input features as the inputs for the model. The model architecture is shown in Fig. 1.

Simulation 2: Analysis of model performance

We trained 37 neural networks (the same number of participants in Korhonen experiments) with the data set comprising the psychoacoustic variables selected in Simulation 1

(see Fig. 1). The average error (for both outputs) of the 37 networks was .05 for the Training set, and .076 for the Test set. These values correspond to 20000 iterations of the training algorithm.

– Insert Figure 1 –

In order to compare the model output with the experimental data for each piece, we calculated the Mutual Information (MI) between the model outputs and the respective target values (experimental data). The MI is a quantity that measures the mutual dependence of the two variables or, in other words, how much they vary together, and it detects both linear and nonlinear correlations between data sets. Because its interpretation, in terms of magnitude, is heavily dependent on data sets used (rendering difficulties for comparisons between different variables), we use a standardized measure for the MI (c.f. Dionísio, Menezes, & Mendes, 2006; Granger & Lin, 1994), based on the global correlation coefficient (λ), defined by

$$\lambda(X,Y) = \sqrt{1 - e^{-2 \cdot I(X,Y)}}$$ [6].

The following analysis will be performed on the network that showed the lowest average RMS error and λ for both data sets (network 24). The RMS errors and λ of each output for all the music pieces are shown in Table 4. Figures 2 and 3 show the Arousal and Valence outputs of the model for Training and Test sets, versus the data obtained experimentally (target values).

– Insert Table 4 –

– Insert Figure 2 –

– Insert Figure 3 –

The model was able to track the general fluctuations in Arousal and Valence for both data sets, although the performance varied from piece to piece. The model performance for Arousal was better for pieces 1, 2, 5, and 6 (RMS_1 = .052, RMS_2 = .040, RMS_5 = .044, and RMS_6 = .052), as shown by the low RMS errors (lower than the mean Arousal for all pieces: RMS_{all} =

.059) and high λ. Pieces 3 and 4 had a higher RMS error than the mean of all the remaining pieces. Nevertheless, only piece 4 shows a λ significantly lower than the remaining pieces). This weaker performance is visible in Figure 3 b). Even though the initial 80s (approximately) of the model predictions show the same increasing tendency of the experimental data, they do not follow the same pattern: they are lower during the initial 50s ("dialogue" between flutes and strings) to which follows a strong increase (only strings playing in bigger number louder) until around 80s of the piece (a transition to a new section in piece).

The best Valence predictions were obtained for pieces 1, 2, 3, and 5 (RMS_1 = .044, RMS_2 = .054, RMS_3 = .045, and RMS_5 = .046): all these pieces had a RMS error lower than the average of all pieces: RMS_{all} = .063). The worst performances were obtained for pieces 4 and 6, although only piece 4 had a λ coefficient significantly lower than the remaining ones (with the exception of piece 5). In these cases, as for the Arousal predictions, poor performance is particularly evident during the initial 80s of the piece, as seen in Figure 3 b).

The successful predictions of the affective dimensions for both known and novel music support the idea that music features contain relevant relationships with emotional appraisals. A visual inspection of the model outputs, confirmed by the RMS and λ measures, also indicates that the model output resembles the experimental data (with the exception of the initial 80s of piece 4). The spatio-temporal relationships learned from the Training set were successfully applied to a new set of stimuli.

These relationships now encoded in the network weights, and the flux of information in the internal (hidden) layer of the neural network represents the dynamics of the internal categorisation (or recombination) of the input stimuli, that enables output predictions. One of the advantages of working with an artificial neural network is the ability to explore the internal

mechanisms that generate the behavior and indirectly show how the model processes the information. In the following paragraphs we will analyze their spatial representation accordingly to Arousal and Valence levels using a method for dimensionality reduction.

Model internal dynamics: discriminant functions. Clustering diagrams of hidden unit activation patterns are good for representing the similarity structure of the representational space. In order to analyze the internal dynamics of our model we use Linear Discriminant Analysis (LDA). The LDA is a classic method of classification using categorical target variables (features that somehow relate or describe the objects). Unlike Principle Component Analysis (PCA), in LDA the groups are known or predetermined[7].

The main purpose of this algorithm is to find the linear combination of features that best separate between classes or object properties. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. Because we are interested in establishing the dynamics of the psychological report, we defined as the classification model the four quadrants of the two-dimentional emotional space (2DES) (Q_1, Q_2, Q_3, and Q_4). We are hypothesizing that the quadrants division of the A/V space represents the underlying internal representations of the model. This method also allows us to identify the hidden units related with each dimension of the categorical space (an important aspects because it will allow for the study of the input-output mapping of the model).

The analysis has shown that two discriminant functions can explain 99.7% of the variance in the data[8]. The canonical correlations of the original data set are .821 for the 1st discriminant function (F_1) and .506 for the 2nd function (F_2). In Fig. 4, we show the two discriminant functions. Each point corresponds to the internal state of the model at a particular

moment in time. The dots color identifies the category hypothesized for each internal state of the model, which correspond to the affective space quadrants (indicated by the labels Q_1 to Q_4).

– Insert Figure 4 –

The model shows an internal discrimination of the input stimuli, which is very similar to the affective space quadrants division. This indicates that the input stimuli were successfully categorized accordingly to their affective value, suggesting that the relationships built in the model transform meaningful patterns of sound features into the Arousal and Valence components of emotion.

As the discriminative power of the model is embedded in the hidden unit activations (the ones that connect to the output), we need to assess the influence of each hidden unit on the pair of canonical variables. This was done by analyzing the factor structure coefficients shown in Table 5. These values correspond to the correlations between the variables in the model and each of the discriminant functions (similar to the factor loadings of the variables on each discriminant function in PCA).

The 1st discriminant function (F_1) receives the highest contributions from H_1, H_3, H_4, and H_5. F_2 receives the strongest contributions from H_2, H_4, and H_5. The next step is to identify how these units relate with the input and output layers. With that information we can estimate the input-output transformations of the model.

– Insert Table 5 –

Input/output transformation: model production rules. To study the relationships between inputs and model predictions it is required an analysis of their relationships with the internal states of the model, which we saw to reorganize the sequence of input stimuli into meaningful affective representations (see previous section). One possibility would be to inspect the weights

matrixes in the model to identify the highest weights. Although simple, this methodology only compares weight values (long-term memory) and excludes the level of activity of each unit (including its bias) and implicit time representations (the short-term memory of the model).

In order to account for the temporal dynamics of the model, the correlations between inputs, hidden, and output units were computed using a Canonical Correlation Analysis (CCA) (Hotelling, 1936). A canonical correlation is the correlation of two canonical variables: one representing a set of independent variables, the other a set of dependent variables. The CCA optimizes the linear correlation between the two canonical variables to be maximized in the context of many-to-many relationships. There may be more than one linear correlation relating the two sets of variables, each representing a different dimension of the relationship, which explain the relation between them. For each dimension it is also possible to assess how strongly it relates each variable in its own set (canonical factor loadings). These are the correlations between the canonical variables and each variable in the original data sets.

In this article the CCA is used to assess the relationships between the sequences of input, hidden and output layers activity. This method permits the analysis of the contribution of each network layer node or (sets of nodes) to the activity of a different layer. Relevant for our analysis are the relationships between input and hidden layers (how the inputs relate with the internal representations of the model), and these with the outputs (which sets of hidden units are more related with the output). In Table 6 we show the details of a CCA for the activity of the neural network layers.

– Insert Table 6 –

Input to hidden: Three canonical variables explain 98.3% of the variance in the data (see left side of Table 6). The first pair of variables loads on P_1, Tx, Ti_1 (inputs set), H_2 and H_5 (hidden

layer). The second, loads only on input D_1, but it loads on all nodes of the hidden layer. The third canonical variable loads on Pv_1, H_2, and H_4. These three dimensions encode the general levels of shared activation in the input and hidden layers.

Hidden to output: Two canonical variables explain all the variance in the data (see right side of Table 6). The first root is correlates strongly with Arousal, and the activity in hidden units H_1 and H_2. The second pair of canonical variables correlates with both Valence (positive) and Arousal (negative), and with the activity in units H_3 to H_5.

Input to output: By taking together these two groups of relationships we can establish qualitative patterns of correlations illustrative of the general model dynamics. Hidden units H_1, H_2, and H_5 have a positive correlation with Arousal. H_5 correlates negatively with Valence and positively with Arousal. H_3 and H_5 correlate negatively with Arousal and positively with Valence. Because Tx, P_1, and Ti_1 relate positively with H_2, they have a positive effect on Arousal. The negative correlation with H_5 indicates that they correlate positively with Valence. D_1 correlated with the activity in all the hidden units. These correlations were consistently positive with Arousal. Finally, Pv_1 shows a negative correlation with Valence (through H_4).

In summary, the general strategies for input-output (sound features -affective dimensions) mapping found are:

Tempo (bpm): fast tempi are related with high Arousal (quadrants 1 and 2), and positive Valence (quadrants 1 and 4). Slow tempi exhibit the opposite pattern;

Texture (multiplicity): thicker textures have positive relationships with Valence and Arousal (quadrants 1, 2, and 4);

Dynamics (loudness): higher loudness relates with positive Arousal;

Mean Pitch (spectrum centroid): the highest pitch passages relate with high Arousal and Valence (quadrants 1, 2, and 4);

Timbre (sharpness): sharpness showed positive associations with Arousal and Valence (especially the first);

Pitch variation (STFT Flux): the average spectral variations relate negatively with Valence and positively with Arousal, indicating that large pitch changes are accompanied by increased intensity and decreased hedonic value.

Discussion and Conclusions

In this paper we presented a novel methodology to study the affective experience of music. From an "emotivist" perspective we considered that music can elicit affective experiences in the listener, focusing on sound features as a source of information about this process. Emotions were represented in terms of two pervasive dimensions of affect: Arousal and Valence. By focusing on continuous measurements of emotion we investigated the relationships between perceptual features of sound and reports of subjective feelings of emotion.

Initially we focused on the reduction of psychoacoustic variables used by Korhonen, in order to identify a group of variables relevant for our hypothesis, but also to reduce the redundancy within the set. The initial simulations allowed us to select 6 variables: dynamics (loudness), pitch level (spectral centroid), pitch variations (mean spectral flux), timbre (sharpness), texture (multiplicity), and tempo. Then we conducted a series of simulations to "tune" and test our model. We used 486 seconds of music (three pieces) as the sample set (used to train the neural network to train the neural network to respond as close as possible to the human participants). A further 632 seconds of music (three pieces) were used as test set. The

model did not have any previous knowledge about these three pieces. We have shown that our models predictions resemble those obtained from human participants.

In terms of modeling technique our model constitutes an advance in several respects. First, we are able to incorporate all music variables together in a single model, which permits to consider interactions among sound features (overcoming some of the drawbacks from previous models Schubert, 1999a). Second, artificial neural networks, as nonlinear models, enlarge the complexity of the relationships between music structure and emotional response observed, since they can operate in higher dimensional spaces (not accessible to linear modeling techniques such as the ones used by Schubert, 1999a and Korhonen, 2004). Third, the excellent generalization performance (prediction of emotional responses for novel music stimuli) validated the model and supported the hypothesis that psychoacoustic features are good predictors of the subjective feeling experience of emotion in music (at least for the affective dimensions considered). Fourth, another advantage of our model is the possibility to analyze its dynamics; an excellent source of information about the rules underlying input/output transformations. This is a limitation inherent in the previous models we wished to address. It is not only important to create a computational model that represents the studied process, but also to analyze the extent to which the relationships built-in are coherent with empirical research. In our analysis we have identified consistent relationships between music features and the emotional response, which support important empirical findings (e.g., Hevner, 1936, Gabrielsson & Juslin, 1996, Scherer & Oshinsky, 1977, Thayer, 1986, Davidson, Scherer, & Goldsmith, 2003; see Schubert, 1999a, and Gabrielsson & Lindström, 2001 for a review).

Our work presented some evidence supporting the "emotivist" views on musical emotions. We have shown that a significant part of the listener's affective response can be

predicted from the psychoacoustic properties of sound. We found that these sound features (to which Meyer referred as "secondary" or "statistical" parameters) encode a large part of the information that allows the approximation of human affective responses to music. Contrary to Meyer's (Meyer, 1956) belief, our results suggest that "primary" parameters (derived from the organization of secondary parameters into higher order relationships with syntactic structure), do not seem to be a necessary condition for the process of emotion to arise (at least in some of its components). This is also coherent with Peretz et al. (1998) study, in which a patient lacking the cognitive capabilities to process the music structure (including Meyer's "primary" parameters), was able to identify the emotional tone of music.

Our current research work focuses on the expansion of the model. In an attempt to overcome the limitations of using a dimensional representation of emotion, we are conducting an experiment, using a similar framework as Schubert (1999a) and Korhonen (2004) but with the additional measurement of physiological activity. We intend to improve the description of the affective experience caused by music by accounting for other components of emotion. Our goal is to assess the relevance of physiological cues for the prediction of the affective experience of music. We are also looking at the identification of individual features in listeners, such as music training/expertise and personality traits, that may alter affective experience. These are also candidates to be incorporated into the model.

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.

Blood, A. J., & 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 , 98*, 11818-11823.

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

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

Brockwell, P. J., & Davis, R. A. (1991). *Time Series: Theory and methods.* (2 ed.). New York, NY, USA: Springer.

Cabrera, D. (2000, July). PsySound 2: Psychoacoustical Software for Macintosh PPC.

Cabrera, D. (1999). PsySound: A computer program for psychoacoustical analysis. *Proceedings of the Australian Acoustical Society Conference*, (pp. 47-54). Melbourne.

Cuddy, L. L., & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics , 57*, 451-462.

Damasio, A. (2000). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness.* London: Vintage.

Davidson, R., Scherer, K., & Goldsmith, H. (2003). *Handbook of Affective Sciences.* Oxford, New York: Oxford University Press.

Dionísio, A., Menezes, R., & Mendes, D. (2006). Entropy-Based Independence Test. *Nonlinear Dynamics , 44*, 351-357.

Dowling, W., & Harwood, D. (1986). *Music cognition.* San Diego (CA), USA: Academic Press.

Elman, J. L. (1991). Distributed Representations, Simple Recurrent Networks, And Grammatical Structure. *Machine Learning , 7*, 195--225.

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

Gabrielsson, A., & Juslin, P. (1996). Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience. *Psychology of Music , 24*, 68.

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. *Music and emotion: theory and research* , 223-248. (P. Juslin, & J. Sloboda, Eds.) Oxford, UK: Oxford University Press.

Granger, C., & Lin, J. (1994). Using the mutual information coefficient to identify lags in nonlinear models. *Journal of Time Series Analysis , 15*, 371-384.

Hevner, K. (1936). Experimental Studies of the Elements of Expression in Music. *The American Journal of Psychology , 48*, 246-268.

Hotelling, H. (1936). Relations between two sets of variables. *Biometrika , 28*, 321-377.

Kivy, P. (1990). *Music Alone: Philosophical Reflections on the Purely Musical Experience.* Ithaca, NY, USA: Cornell University Press.

Korhonen, M. (2004). *Modeling Continuous Emotional Appraisals of Music Using System Identification.* Master's Thesis, University of Waterloo.

Kremer, S. (2001). Spatiotemporal Connectionist Networks: A Taxonomy and Review. *Neural Computation , 13*, 249-306.

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception , 13*, 401-432.

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

Krumhansl, C. L. (1991). Melodic structure: Theoretical and empirical descriptions. *Music, language, speech and brain* . (J. Sundberg, Ed.) London, UK: MacMillian.

Langer, S. K. (1942). *Philosophy in a new key.* Cambridge, MA, USA: Harvard University Press.

Ljung, L. (1987). *System identification: theory for the user* (2 ed.). Old Tappan, NJ, USA: Prentice Hall.

Malloch, S. (1997). *Timbre and technology: an analytical partnership. The development of an analytical technique and its application to music by Lutosławski and Ligeti.* PhD Thesis, University of Edinburgh, Edinburgh.

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

Narmour, E. (1992). *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model.* Chicago, IL, USA: University Of Chicago Press.

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

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

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

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

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

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

Russell, J. A. (1989). Measures of emotion. *Emotion: Theory, research, and experience , 4* . (R. Plutchik, & H. Kellerman, Eds.) Toronto: Academic.

Scherer, K. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research , 33*, 239-251.

Scherer, K., & Oshinsky, J. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion , 1*, 331-346.

Schubert, E. (2001). Continuous Measurement of Self-Report Emotional response to music. *Music and emotion: theory and research* , 393-414. Oxford, UK: Oxford University Press.

Schubert, E. (1999a). *Measurement and time series analysis of emotion in music.* PhD Thesis, Univ. of New South Wales.

Schubert, E. (1999b). Measuring emotion continuously: Validity and reliability of the two dimensional emotion space. *Australian Journal of Psychology , 51*, 154-165.

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

Sloboda, J., & Lehmann, A. (2001). Tracking Performance Correlates of Changes in Perceived Intensity of Emotion During Different Interpretations of a Chopin Piano Prelude. *Music Perception , 19*, 87-120.

Thayer, J. (1986). *Multiple Indicators of Affective Response to Music.* PhD Thesis, New York University, Graduate School of Arts and Science.

Tzanetakis, G., & Cook, P. (1999). MARSYAS: a framework for audio analysis. *Organised Sound*, *4*, 169-175.

Wundt, W. (1896). *Grundriss der Psychologie (Outlines of Psychology).* Leipzig: Engelmann.

Zwicker, E., & Fastl, H. (1990). *Psychoacoustics.* New York, NY, USA: Springer.

Author Note

The authors would like to acknowledge the courtesy of Mark Korhonen for sharing his experimental data, and the financial support from the Portuguese Foundation for Science and Technology (FCT). We are also grateful to Dr. Leonid Perlovsky for his relevant comments and suggestions, Dr. Stephen Hanson and Dr. Andreia Dionísio for their support and discussion on the analysis methodology, and to Andrew Hennell for his help on the preparation of this manuscript.

Notes

[1]In the screening tests, used to test I.R.s ability to process music, she did not give any indication on that she could perceive and/or interpret pitch and temporal variations in melodies.

[2]Data available online at http://www.sauna.org/kiulu/emotion.html, courtesy of the author.

[3]Although these algorithms are not specific measures of melodic contour, they have been successfully used as such in music information retrieval applications (Korhonen, 2004). Nevertheless, in this article we refer to this variable as pitch variation because it characterizes better the nature of the encoding. Moreover, the relationships between pitch variations and emotion were the object of some studies (e.g., Scherer & Oshinsky, 1977), as described in Schubert (1999a).

[4]T and Tx were chosen as the variables for the initial features for a few reasons. First, is that they are the only variable for the sound features that they represent. A second important factor is that T and Tx are expected to contain important information about changes in the affective experience (Schubert, 1999a).

[5]In the tables we indicate no index when we include all variables from that music feature; in this case D indicates D_1 and D_2.

[6]X and Y are the data sets being compared and $I(X,Y)$ is the MI score.

[7]Both methods are very similar because they look for linear combinations of variables which best explain the data; the essential difference consists of the rules for classification (clustering), which is based on distance measures in PCA while LDA explicitly attempts to model the difference between the classes.

[8]This does not mean that we can reduce the number of units in the model, but instead that some of these units might vary along similar dimensions. As we'll see all the hidden units have relevant contributions to at least one of the discriminant functions.

Table 1

Pieces used in Korhonen's experiment and their aliases for reference in this paper. The pieces were taken from the Naxos's "Discover the Classics" CD 8.550035-36

Piece ID	Alias	Title and Composer	Duration	Set
1	Aranjuez	Concierto de Aranjuez -II. Adagio (J. Rodrigo)	165s	Training
2	Fanfare	Fanfare for the Common Man (A. Copland)	170s	Training
3	Moonlight	Moonlight Sonata -I. Adagio Sostenuto (L. Beethoven)	153s	Test
4	Morning	Peer Gynt Suite No 1 -I. Morning mood (E. Grieg)	164s	Training
5	Pizzicato	Pizzicato Polka (J. Strauss)	151s	Test
6	Allegro	Piano Concerto no.1 -I. Allegro maestoso (F. Liszt)	315s	Test

Table 2

Psychoacoustic variables considered for this study. The aliases indicated will be used in this article to refer to the variables in this table.

Musical Property	Musical Feature	Alias
Loudness Level	Dynamics	D_1
Short Term Maximum Loudness	Dynamics	D_2
Power Spectrum Centroid	Mean Pitch	P_1
Mean STFT Centroid	Mean Pitch	P_2
Mean STFT Flux	Pitch Variation	Pv_1
Standard Deviation STFT Centroid	Pitch Variation	Pv_2
Standard Deviation STFT Flux	Pitch Variation	Pv_3
Sharpness (Zwicker and Fastl)	Timbre	Ti_1
Timbral Width	Timbre	Ti_2
Mean STFT Rolloff	Timbre	Ti_3
Standard Deviation STFT Rolloff	Timbre	Ti_4
Beats per Minute	Tempo	T
Multiplicity	Texture	Tx

Table 3

RMS error for each input data set using a model with 5 hidden units. The values shown were averaged across 3 simulations for each test case.

Input Set	RMS Train		RMS Test		Mean RMS
	Arousal	Valence	Arousal	Valence	
T-Tx-D	.056	.061	.068	.080	.066
T-Tx-D_1	.058	.058	.072	.081	.067
T-Tx-D_2	.066	.056	.088	.077	.072
T-Tx-Ti	.074	.067	.088	.087	.079
T-Tx-Ti_1	.069	.063	.082	.093	.077
T-Tx-Ti_2	.105	.073	.098	.080	.089
T-Tx-Ti_3	.108	.074	.135	.121	.110
T-Tx-Ti_4	.110	.076	.130	.087	.101
T-Tx-P	.080	.072	.106	.095	.088
T-Tx-P_1	.072	.066	.107	.083	.082
T-Tx-P_2	.136	.083	.233	.106	.140
T-Tx-Pv	.100	.062	.119	.083	.091
T-Tx-Pv_1	.101	.064	.130	.086	.095
T-Tx-Pv_2	.108	.070	.134	.083	.099
T-Tx-Pv_3	.102	.067	.133	.090	.098
T-Tx-D_1 -P_1 –Ti_1 -Pv_1	.048	.049	.068	.076	.060

Table 4

Comparison between the model outputs and experimental data: root mean square (RMS) error and global correlation coefficient (λ).

Piece	RMS error		MI (λ)		Set
	Arousal	Valence	Arousal	Valence	
1	.052	.044	.940	.770	Training
2	.040	.054	.761	.874	Training
3	.061	.045	.754	.652	Test
4	.085	.081	.539	.556	Test
5	.044	.046	.899	.490	Training
6	.052	.082	.961	.736	Test

Table 5

Factor Structure Matrix: correlations between discriminant variables and each hidden unit.

Hidden unit	F_1	F_2
H_1	-.489	-.246
H_2	.371	.896
H_3	-.788	.291
H_4	-.520	.569
H_5	.633	-.604

Table 6

Canonical Correlation Analysis (CCA): the canonical correlations (the canonical correlations are interpreted in the same way as the Pearson's linear correlation coefficient) quantify the strength of the relationships between the extracted canonical variates to assess the significance of the relationship. To assess the relationship between the original variables (inputs and hidden units activity) and the canonical variables, we also include the canonical loadings (the correlations between the canonical variates and the variables in each set)

Canonical Loadings (Input/Hidden)			
Variable	var. 1	var. 2	var. 3
H_1	-.398	-.633	-.028
H_2	.479	.657	-.437
H_3	.144	-.891	-.238
H_4	.159	-.647	-.632
H_5	-.637	.645	.018
T	.264	.478	.151
Tx	.608	.280	.217
D_I	.450	.674	.139
P_I	.819	.297	.432
Ti_I	.748	.420	.262
Pv_I	.187	.270	.825
Canon Cor.	.725	.546	.448
Pct.	61.1%	23.4%	13.8%
Wilks' L.	0.259	0.545	0.777
Sig.	.000	.000	.000

Canonical Loadings (Hidden/Output)		
Variable	var. 1	var. 2
H_1	-.504	.482
H_2	.978	-.055
H_3	-.291	.862
H_4	.014	.797
H_5	-.074	-.973
A	.765	-.644
V	.260	.966
Canon Cor.	.987	.984
Pct.	56.0%	44.0%
Wilks' L.	0.001	0.032
Sig.	.001	.000

Figure Captions

Figure 1. Neural network architecture and units identification (model used in simulations)

Figure 2. Training data set (Aranjuez, Fanfare and Pizzicato): Arousal and Valence model outputs compared with experimental data

Figure 3. Test data set (Moonlight, Morning and Allegro): Arousal and Valence model outputs compared with experimental data

Figure 4. Canonical Discriminant Functions plot: each point corresponds to the internal state of the model at a particular moment in time. The dots color identifies the internal states of the model belonging to each of the categories hypothesized (the affective space quadrants), and the labels (Q_1 to Q_4) indicate the correspondent quadrant in the 2DES to which each color group belongs to.

Figure 1

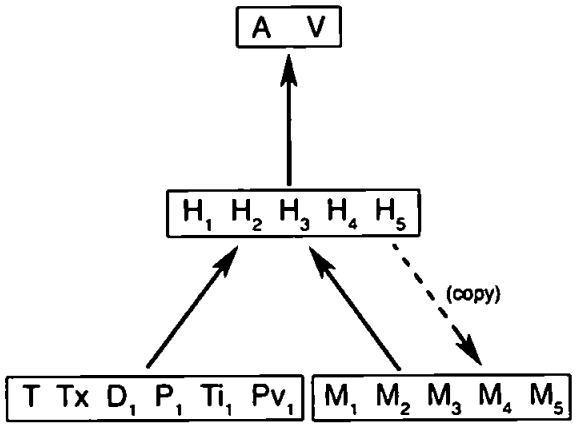


Figure 2

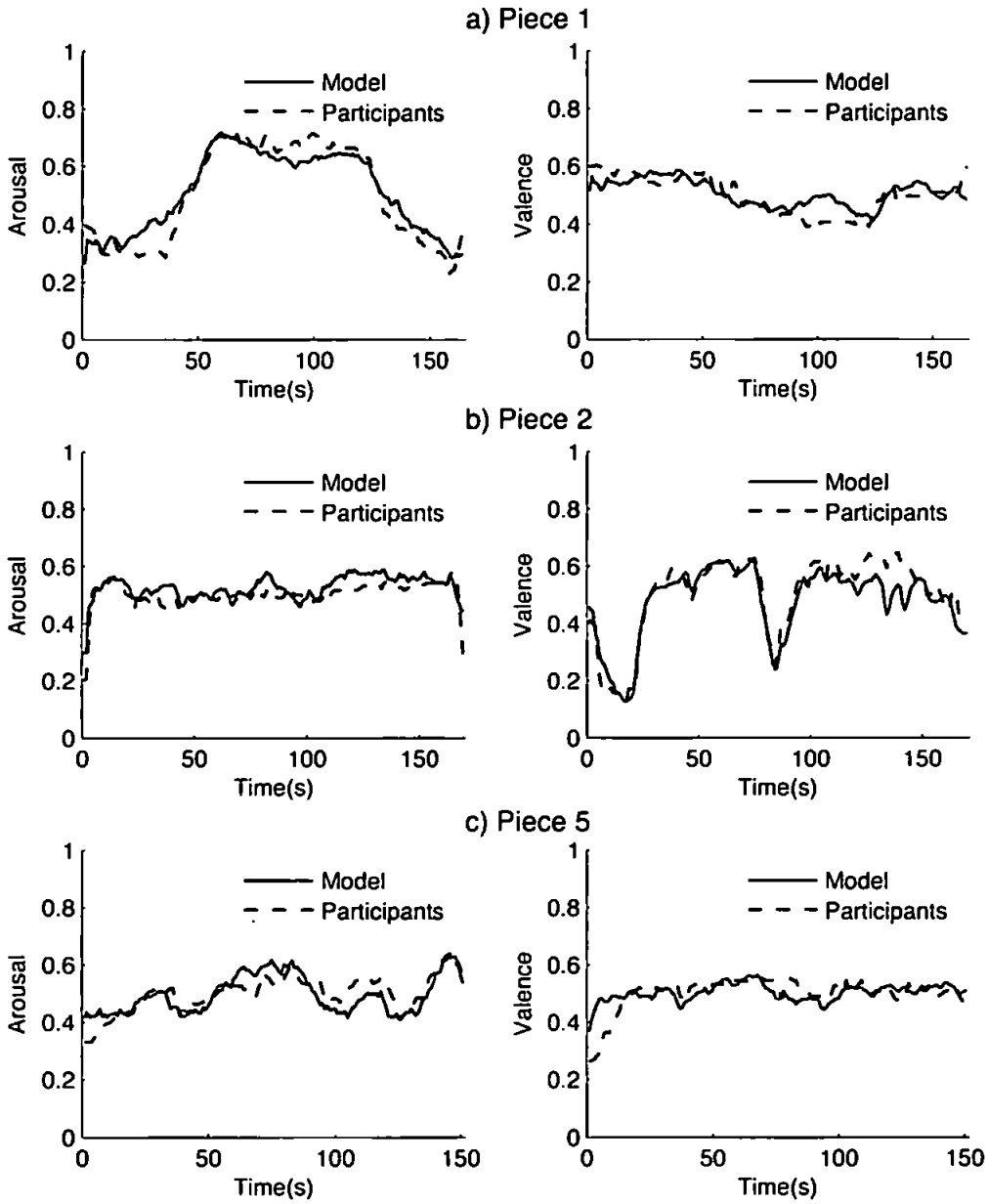

a) Piece 1
b) Piece 2
c) Piece 5
Model
Participants
Arousal
Valence
Time(s)
0
0.2
0.4
0.6
0.8
1
50
100
150

Figure 3

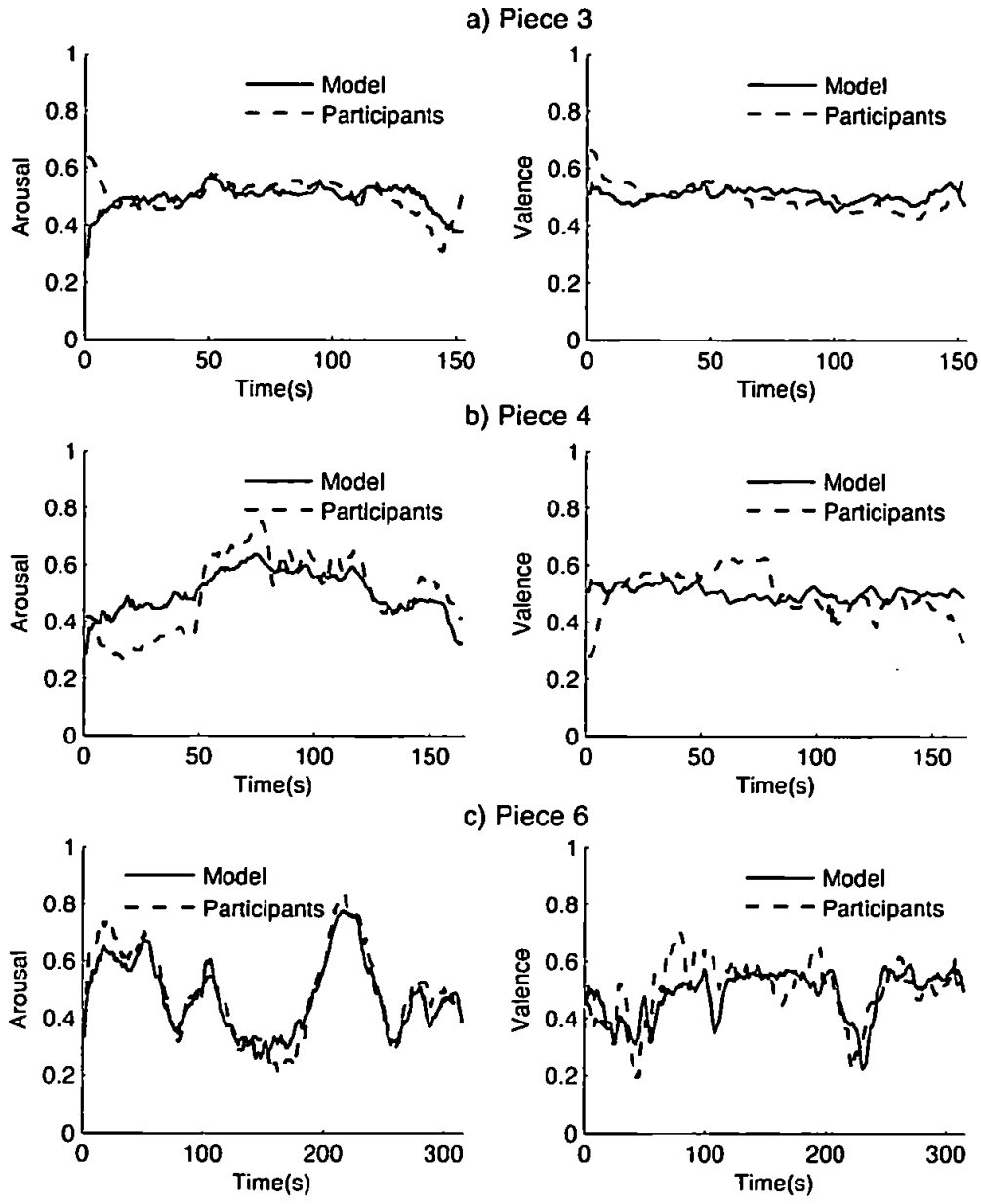
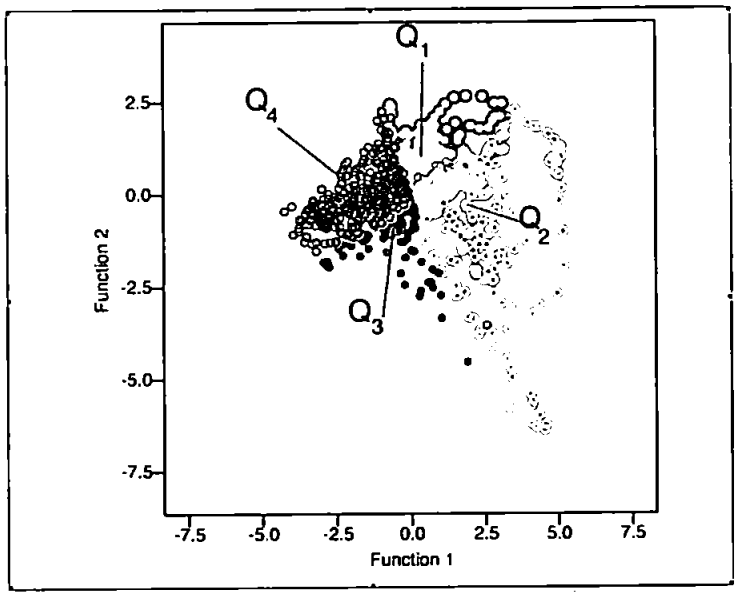

a) Piece 3
b) Piece 4
c) Piece 6
Model
Participants
Arousal
Valence
Time(s)
0
0.2
0.4
0.6
0.8
1
50
100
150
200
300

Figure 4


Q1
Q4
Q2
Q3
2.5
0.0
-2.5
-5.0
-7.5
Function 2
-7.5
-5.0
-2.5
0.0
2.5
5.0
7.5
Function 1

Coutinho, E. & Cangelosi, A. (2008). Psycho-physiological patterns of musical emotions and their relation with music structure. *In Miyazaki, K., Hiraga, Y., Adachi, M., Nakajima, Y. & Tsuzaki, M. (Eds.), Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC10).* Sapporo (Japan).

Psycho-Physiological Patterns of Musical Emotions and Their Relation With Music Structure

Eduardo Coutinho,[*1] Angelo Cangelosi [*]

[*] *Adaptive Behaviour and Cognition Research Group, SoCCE, University of Plymouth, United Kingdom*
[1] eduardo.coutinho@plymouth.ac.uk

This study investigates the dynamics of psychological and physiological reactions during music listening to determine whether differentiated psychological and physiological patterns could be related with differentiated patterns of music variables.

We asked 39 participants to give continuous self-reports of the intensity (Arousal) and hedonic value (Valence) of the emotions felt while listening to 9 pieces of western instrumental art (classical) music. Simultaneously we recorded their Heart Rate (HR) and Skin Conductance Response (SCR) for the full length of the pieces.

Arousal was found to increase for higher levels of tempo, loudness, mean pitch, sharpness and timbral width, and Valence was correlated with variations in tempo and (tonal) dissonance. Psychological and physiological reports showed that increased Arousal and Valence are related with increased SCR, while increased HR related with higher Arousal. We also found a negative relationship between Valence and Heart Rate.

Our study supports the idea that significant patterns of interactions between music structure and differentiated levels of intensity and hedonic value do exist in musical emotions. Our study also shows differentiable physiological patterns across different emotions, supporting the claim that physiological arousal is also a component of musical emotions. We are currently creating a neural network model to analyse in detail such interactions.

Coutinho E. & Cangelosi A. (2007). Modeling emotion and embodiment in multi-agent systems. *In H. Hexmoor & C. Thompson (Eds.), Proceedings of 2007 International Conference on Integration of Knowledge Intensive Multi-Agent Systems (KIMAS07).* Waltham (MA, USA): IEEE Press, pp. 133-138.

Emotion and Embodiment in Cognitive Agents: from Instincts to Music

Eduardo Coutinho and Angelo Cangelosi
Adaptive Behaviour and Cognition Research Group - School of Computing, Communications & Electronics
University of Plymouth
Drake Circus, Plymouth PL4 8AA, United Kingdom
{eduardo.coutinho , angelo.cangelosi}@ plymouth.ac.uk

Abstract - This paper suggests the use of modeling techniques to tack into the emotion/cognition paradigm. We presented two possible frameworks focusing on the embodiment basis of emotions. The first one explores the emergence of emotion mechanisms, by establishing the primary conditions of survival and exploring the basic roots of emotional systems. These simulations show the emergence of a stable motivational system with emotional contexts resulting from dynamical categorization of objects in the environment, in answer to survival pressures and homeostatic processes. The second framework uses music as a source of information about the mechanism of emotion and we propose a model based on recurrent connectionist architectures for the prediction of emotional states in response to music experience. Results demonstrate that there are strong relationships between arousal reports and music psychoacoustics, such as tempo and dynamics. Finally we discuss future directions of research on emotions based on cognitive agents and mathematical models.

1. INTRODUCTION

Recent findings in neurosciences, psychology and cognitive sciences indicate the surprising role of Emotions in intelligent behavior. Particularly interesting for us are the studies looking at physiological interferences, and the relation between body and affective states, as well to the evolutionary mechanisms. Emotions have an important role in behavior and adaptation in biological systems.

This idea has recently gained special attention in computational models of cognition and behavior (e.g. [1,2]), following new theoretical approaches to cognition based on embodiment theories [3]. Whilst some of these models focus on different properties of an emotional system for task solving issues (e.g. using facial expressions for social engagement), we are interested in using computational models to understand the basic mechanisms of the emotional systems.

We are developing computational models of simulated autonomous agents that use emotion as a mechanism for organization of behavior. This way we intend to create an integrated model of instincts, perception, motivation and action, based on artificial environments and embodiments. In our modeling approach we share the neurobiological and evolutionary perspectives to Emotions, as discussed in the following sections.

We suggest that the agent should be embodied so as to allow its behavior to be affected by motivational processes, focusing on the internal demands and activity. By artificial embodiment we mean that the agent has a virtual physical body whose states can be sensed by the agent itself. We will discuss a theoretical framework and specific scenarios to test our hypothesis. We present results for one test framework on the emergence of motivational processes in embodied agents. The preliminary framework seems to be effective and versatile enough to allow the agent to adapt itself to unknown world configurations, maintaining controlled healthy states. We demonstrate that body/world categorizations and body maps can evolve from the simple self-survival rule. The results are coherent with Antonio Damasio's definition of background emotional system [4].

We are also studying the extension of this model to musical emotions [5]. In this paper we will also present experimental results derived from the application of recurrent connectionist architectures for the prediction of emotional states in response to music experience. Results demonstrate that there are strong relationships between arousal reports and music psychoacoustics. We believe that this approach can also contribute for different cognitive studies, and its large spectrum of applications may be specially interesting in computational approaches for intelligent systems. A potential field of application of these models arises from recent theoretical proposal in robotics, more specifically that of Internal Robotics [6].

2. EMBODIED EMOTIONS

The study of Emotion is increasingly considered of paramount importance for modern sciences. Throughout the last century emotion has appeared as a main research topic in several areas of knowledge, specially after the emergence of post-Cartesian theories exposing the

1-4244-0945-4/07/$25.00 ©2007 IEEE

relevance of emotional systems at all levels of behavior. Charles Darwin, William James, Walter Cannon, Wilhelm Wundt, Susanne Langer, among others, introduced great developments in the study of emotions, and their ideas are still largely considered as references for many emotions theorists. Several are the definitions for the process of Emotion, and consequent research focus, although current research generally accepts that it can be described as a multi-modal mechanism, with several processes involved. including appraisal, basic emotions, physiological responses, and subjective feeling states [4,7,8].

It is important to stress that the process of emotion differs in important aspects from other psychological processes. For instance, emotion is an embodied experience with specific behavioral patterns (facial expressions, autonomic arousal, etc.); it is less susceptible to our control and also expressed at the unconscious level; has the capacity to affect other cognitive processes (e.g. decision making), though not only confined to the old sub-cortical structures in our brains [4]. Damasio suggests that the processes of emotion and feeling are part of the neural machinery for biological regulation, whose core is formed by homeostatic controls, drives and instincts.

Emotions are complicated collections of chemical and neural responses, organized in various patterns; all emotions have some regulatory role to play, leading in one way or another to the creation of circumstances advantageous to the organism exhibiting the phenomenon. The biological function of emotions can be divided in two: the production of a specific reaction to the inducing situation (e.g. run away in the presence of danger), and the regulation of the internal state of the organism such that it can be prepared for the specific reaction (e.g. increased blood flow to the arteries in the legs so that muscles receive extra oxygen and glucose, in order to escape faster). Emotions are also inseparable from the idea of reward or punishment, of pleasure or pain, of approach or withdrawal, of personal advantage or disadvantage.

Obviously this process is dependent on the mechanisms of integration that relates the inducer with the organisms internal state. It is on this border that the outer world of objects is separated from the inner world of emotions and concepts. Instincts continuously drive behavior towards the self protection and homeostatic balance. By evaluating concepts derived from internal and/or external stimuli (the internal model of the outer world), the process of emotion arises, integrating both dimensions and modulating ongoing processes in the brain. Survival mechanisms are related this way to emotions and feelings (the synthesis of emotion), in the sense that the they are regulated by the same mechanisms. In this line the internal representations of the outer world can induce emotional states, by interacting with both body and psyche.

Emotion as Arousal

The autonomic nervous system (ANS) regulates the body and its readiness for action. The correspondent physiological variations are referred as changes in arousal. Changes in arousal are related to emotional states or experiences, having both brain and physiological inducers. Regarding specifically the role of physiological responses on the emotional experience and elicitation (and the interaction with the brain), there have been several models proposed. As discussed in [9] regarding the role of the body in the emotional experience, three main models of emotion can be distinguished: (i) the undifferentiated arousal model, (ii) the cognitive appraisal model, and (iii) the central network model. The main idea behind the first model (see [10]) is that body responses increase with emotional intensity, but their pattern is not differentiated across the different emotional states. In this line cognitive information and/or the specific context differentiate the type of emotion, while bodily activation (arousal) determines the intensity of that emotion. One practical prediction of this model is that the perception of the emotional intensity can be influenced by the arousal intensity. The main finding of this research has been the fact that after the exposure to an arousing stimulus, the following emotional feeling state is intensified. This phenomenon is called activation transfer [11]. The second model focuses on the body changes as a function of cognitive appraisal processes [8], or action readiness. In this line of research, the patterns of body changes are the combined result of the several cognitive appraisal components, although the fact that the body itself might generate emotional states is quite marginalized in this model. Finally, from the third model perspective, emotions share different neural and cognitive mechanisms and pathways, and their pattern of interaction defines the emotional nature. In short, the patterns of body changes are differentiable across emotions. The activation of the body with a pattern related to a specific emotion will, in certain conditions, elicit that emotion (the peripheral feedback). This last process is automatic at an implicit level [4].

In summary, the underlying idea common to all these mechanisms is that a specific emotion can be elicited by creating specific body state patterns (by manipulating the body), even outside the awareness of the individual. An event (appraised via cortical or subcortical routes) elicits physiological changes that facilitate action and expressive behavior. These changes are accompanied by, and contribute to, an affective feeling state. Motoric and visceral feedback can contribute to the intensity and hedonic value of an emotional experience: consciously or subconsciously, individuals use their body state as a clue to the valence and intensity of the emotion they feel.

These ideas give rise to an interesting discussion about interaction between affect and logic, emotions and cognition. Sustained for an evolutionary perspective certain organizational principles in the brain might reflect emotional states (for further details and references on emotions, cognition and behavior, please refer to [4,9]). Levine [14] as suggested that the ancient emotional centers are also involved in the mediation of the mind-

body connection of aesthetic emotions. Furthermore, Perlovsky [15] suggested that aesthetic emotions are related to cognition, to improvement of knowledge and are not directly related to specific bodily needs and experiences. Aesthetic emotions mediate value-relationships among cognitive concepts. A large number of these emotions is needed to unify knowledge in its relationship to our cognition as a whole. Music creates this diversity of emotions.

In the following section we will present two possible frameworks for the study of these issues. One focus on the biological subtract of artificial embodied agents, endowing them with basic instincts and exploring the role of emotional mechanisms in an adaptive task; the other investigates the nature of Music stimuli as a source of affect, in order to understand the nature of the signals that might interfere with arousal mechanisms, which have impact at all levels of behavior.

3. FRAMEWORKS

Experiment I: Instincts and Action

We created a conceptual A-Life model to implement artificial worlds inhabited by autonomous emotional agents, modelling the agent based on biologically plausible principles. We focused on the idea of having an embodiment (in the sense that the agent has a virtual physical body whose states can be sensed by the agent itself) so that low level tasks (e.g. satiate body needs) influence its overall performance, by affecting its behavior. A neural network endows the agent with cognitive capabilities, processing information related to its body, and to its environment. The agent's emotional state is mirrored into a set Background Emotions. This term is used by Damasio [4] for the responses caused by "...certain conditions of internal state engendered by ongoing physiological processes or by the organism's interactions with the environment or both" [1]. Note that in this paper we consider these emotions to be derived from Instincts.

The simulation environment (programmed in C++) is represented as a two-dimensional world populated by several objects. An autonomous agent inhabits this artificial world and is able to move within the borders that define its limits. Each object has different representations corresponding to a physiological interference. For instance a red object correponds to a source of food, though, simplisticly, a resource of blood sugar for the agent. When exhibiting the will to eat, by interacting with this object, the agents blood sugar level is increased. Similar objects were created (using different colors) related to physiologic demands related to energy, endorphine, and vascular volume. We also created obstacles in the world that, in the simplest case, are only the topological limits (or borders) of the world; these borders are considered as sources of pain.

Currently, the agent comprises three main formalized systems: Perceptual, Nervous and Motor.

In order to perceive the world, an agent contains a retina (represented as a color array) that resembles a biological retina on a functional level. It senses a bitmap world (environment) through a ray tracing algorithm, and it incorporates the attenuation of visual cues in function of the distance of the objects in relation to the agent. This information is feeded to the nervous system.

The nervous system includes a feed-forward neural network (NN) with a genetically encoded structure (fixed during lifetime). The neural network is organized in layers: an input layer (two groups: retina and body sensors), an output layer (two groups: instincts and motor control) and a hidden layer (with excitatory-only and inhibitory-only neurons). The instincts drive the agent attention towards specific needs. They are only controlled by the agent embodiment, which reacts to its environment, with no other interference, and they translate physiological changes into specific alarms or urges to action (e.g hunger if blood sugar is low). In any moment the agent can be hungry or satiated, tired or energetic, etc. To express its desire to act in the environment, the agent possesses a set of Motivations, which correspond to the level of will to adopt a certain behavior (eat, drink, etc.). The Motivational System is controlled by the neural process.

The agent also controls a motor system through linear and angular speed signals, allowing it to travel around the world (including obstacle avoidance and object interaction). These signals are provided by the neural network, which means that motor skills also have to be learned. With this capabilities the agent will be able to navigate in its environment, approaching or avoiding certain states.

The agent learns through a reward and punishment algorithm, in order to adapt itself to the environment by interacting with it. Our algorithm is inspired in Rolls' "Stimulus-Reinforcement Association Learning" [13].

The aim of this exercise is to design an embodied agent-based cognitive model and establish how an emotional system can emerge from self-regulatory Homeostatic Processes. The objective is to understand the role and the importance of Emotions in self-survival tasks; hence one of the reasons to implement a single-agent system at this stage. We are also interested in studying how the regulation of the Homeostatic Processes can influence world categorization and decision making (currently at a low level and for single tasks). We also analyze how emotions act as a system of internal rewards, that preserve the system, and permit a continuous adaptation process in self-survival tasks, by signalling and scaling pleasant or unpleasant interactions. Detailed simulations and results were reported in [12]. Here we give an overview of the main achievements.

With this framework we aim, at this point, to test three main hypothesis: (i) an emotional system can emerge from the interaction of self-regulatory Homeostatic Processes and the Environment; (ii) the regulation of the

1 Page 52

Homeostatic Processes influences world categorization and decision making by attributing hedonic values (or relations with Insticts) to objects, and by affecting cognitive processes (e.g. driving attention); (iii) emotions act as a system of internal rewards, that preserve the system, and permit continuous adaptation in self-survival tasks, by signalling and scaling pleasant or unpleasant interactions/stimuli.

Results overview - The evolution of the Fitness value in time showed an overall increase of the agent's ability to regulate its body state. The agent is not only capable of increasing its Fitness, but it does so by maintaining a "healthy behavior". Extreme body states were avoided, showing the ability of the agent to regulate its own body status, by coping with its metabolism and managing competitive internal stimuli.

We analysed the system further in order to better understand the dynamics of the NN and the learning process. Particularly, we wanted to understand how the agent categorizes the stimuli and how this information is integrated giving rise to behavioral changes. The first 2 Principle Components of a PCA analysis of the hidden units activations (when presenting to the agent all the objects isolatedelly and one at the time) showed that the agent was able to categorize the world. In fact, identical external stimuli (objects) are represented internally in a specific and dedicated way. In another test scenario we varied the physiological state of the agent, and we found that there is also a clear definition of the different body states.

Summarizing we hypothised that an emotional system, as a process of integration, would emerge from the interaction of self-regulatory Homeostatic Processes and the Environment. Moreover it would affect other cognitive processes namely related with internal concepts of the world. In our experiment we found that the agent is able to identify its own body needs and attribute dynamical meanings to the objects. This was specially evident by the complete separation of the different states of well-being (over-stimulated, homeostatic regime or under-stimulated) when stimulated by the world objects. These results confirm our hypothesis.

Experiment II: Differentiation and Synthesis

As emphasized by Perlovsky [15], Music might well be the main mechanism of differentiation of emotions. For that it is also considered as a mechanism of synthesis, since it can interact with the entire wealth of human experience. Music is able to speak to both conscious and subconscious states of mind, sometimes in unique ways, which have special impact on our lives for its ability to create a connection between the outer world of objects and stimulus, and the inner world of emotions.

While in the previous approach we hypothesized a scenario to observe the influence of instincts and emotion on behavior and adaptation, we now suggest to analyze the mathematics of musical stimulus as a source of information about systems of differentiation and synthesis. We will develop now our ideas about a possible line of investigation, which seaks for mathematical relantionships between acoustic properties of sound and psycho-physiological reports of emotion.

Music and Emotions - Music offers an extraordinary platform for the study of Emotion. People often report that their primary engagement with music is emotional, and there is a wide acceptance that musical stimuli are among the most powerful triggers of strong emotions [16]. But are these musical emotions similar to the emotions referred previously? Do they contribute for an individual's well being? Do they have any psychological consequences? To a certain extent recent research suggests a positive answer to these questions, particularly when looking at the interaction of the temporal patterns of music with body and brain.

Current research supports the hypothesis that music perception is a distributed process within the brain and is involved in an interactive spatiotemporal system of different neural networks [17,18], including even neural networks specialized for other cognitive processes (e.g. emotion, memory, language). Experimental evidence suggests the participation of both hemispheres in music perception, although it is possible to differentiate some interesting specializations [19]. For instance the left hemisphere seems to be related to perception of timing and rhythm, while the right hemisphere specializes in pitch and timbre perception. Along with the spatiotemporal patterns of neural activity presented to the primary auditory system, the brain engages in other processes, for instance in the motor system, language, emotion and reward related areas. Focusing on the emotion dimension, some researchers [7,20,21] suggest that music derives its affective power from dynamic aspects of the brain systems, which control emotional processes but are distinct, although interactive, with other cognitive processes. This hypothesis follows Susanne Langers ideas about the existence of shared properties in patterns of physical and mental states, emotion, and music.

Support for these ideas comes, for instance, from research with brain damage patients. It has been shown that the emotional appreciation of music can be maintained even in the presence of severe perceptual and memorization deficits, though reinforcing the idea that sub-cortical mediation is involved in emotional judgments [22]. Due to these interactions certain basic mechanisms related to motivation/emotion in the brain can be elicited by music. This gives rise to changes in the body and brain dynamics, and to interferences with ongoing mental and bodily processes. This multi-modal integration of musical and non-musical information might take place in the brain, opening a window for associations between music, emotion, and physiological states. Past research also indicates the existence of a possible relationship in the dynamics of musical emotion and the cognition of musical structure.

We suggest that the study of the music dimensions can unveil important information about the configuration of the stimuli that create specific, unique, and relevant

patterns of mental and bodily activity. We are looking at the spatiotemporal relationships between music psychoacoustics and levels of psychological and physiological arousal. In [5] we focused on the contribution of mathematical and computational modelling techniques for understanding the relationship between music elements and evoked emotions. In particular, we support the use of spatiotemporal models. We applied the use of the connectionist paradigm to investigate the relationship between music (psychoacoustic features) and emotional responses (quantified as the arousal level).

Experiment overview - We focus the present analysis to psychoacoustic data and psychological arousal, aiming to test our hypothesis on emotional engagement during music listening. To do so we used Korhonen's experimental data, made available online in the researchers website [23]. Korhonen used 6 pieces of classical music for his experiments. Volunteers used a continuous arousal and valence scales (with a sample rate of 1s), to rate the emotion thought to be expressed by the music. The musics used were: Music 1 - Concierto de Aranjuez Adagio (Rodrigo); Music 2 - Fanfare for the Common Man (Copland); Music 3 – Moonlight Sonata Adagio Sostenuto (Beethoven); Music 4 - Peer Gynt Morning (Grieg); Music 5 - Pizzicato Polka (Strauss); Music 6 - Piano Concerto No. 1 Allegro Maestoso (Liszt). For our experiments we use all the 18 psychoacoustic variables chosen by Korhonen to encode the musical input to the system, since we want to include a wide spectrum of information about the musics. Arousal corresponds to the single output of the network. Music 1 to Music 5 were used to train the system, while Music 6 was used to test the generalization response to novel stimuli. All simulations are divided into two phases: the first corresponds to the training phase, and the second to the tests phase. During the former inputs are presented sequentially to the network input layer, where the training algorithm combines the current input with the previous activation of the hidden layer (through the previous states history stored in the context units) and activates the hidden units with the combined input. The output derived from propagating the inputs to the output is then compared with the reference values correspondent to the desired output (in this case the values of arousal obtained from the experiments correspondent to the current musical input). The MSE error calculated from this comparison is then used to adjust the values of the weights of all the trainable connections (which are all except the links between hidden and context layers, which are maintained constant to 1.0), in order to move the network outputs closer to the desired targets (again the values for arousal or valence were obtained experimentally).

Results overview - Results show that there are strong mathematical relationships between music psychoacoustics and arousal reports, having the model explained 85.0 % of the output variation , using the music inputs as predictors. We identify that there are not only spatial relationships between music variables, but also temporal. Arousal predictions derive from information related with different musical dimensions at different stages in the piece of music. Tempo, Dynamics, Mean Pitch and Texture were the variables which contributed with more information to the system. The resultant model is now being analyzed in order to establish the mathematical relationships that underly the interaction between music and psychological arousal. These results agree with research in music and emotion, as extensively reviewed by [24]. In this work strong relationship between arousal and dynamics are reported, as well as between arousal and tempo. Schubert also observes that his results fail to confirm his thesis in what regards to other psychoacoustic variables. In [5] we report that, apart from tempo and loudness, also mean pitch and texture, can explain great part of the variations in arousal ratings. These results improve significantly previous models.

We are able to demonstrate that a spatiotemporal connectionist model trained on music and arousal self-report data is capable of representing the process and generalizing the level of arousal in response to novel music input. The model is also capable of identifying the main variables responsible for such an emotional rating. Ongoing work aims to provide a framework for the understanding of the relationship between musical features and perceived emotions, based on the ideas we presented in [5]. Implications from this research might allow for the mathematical modeling of arousal variations during music listening, opening a way for their incorporation in artificial agents, allowing for its use in scenarios like the one we presented in the previous section.

3. CONCLUSIONS

We suggested the use of modeling techniques to tack into the emotion/cognition paradigm. For that we presented two possible frameworks that can account for such investigation. One explores the emergence of emotion mechanisms, by establishing the primary conditions of survival and exploring the basic roots of emotional systems; the other uses music as a source of information about the mechanism of emotion. We propose to use these frameworks as a basis to develop a simulation scenario to study the dynamics of emotion and its relation with cognition. Recent advances in neuroscience and other studies of emotion, endow us with new information that can be successfully applied in modeling frameworks.

We addressed the notion of the emergence of a stable emotional system by means of self-regulatory Homeostatic Processes, and we demonstrated that it is possible to model such phenomenon. As suggested by Damasio [4], environmental events of value should be susceptible to preferential perceptual processing regarding their pleasant or unpleasant meaning. We believe that the architecture and particularly the reward system (the agent's appetite for well-being) were responsible for the

emergence of stable emotional systems in our simulations. Furthermore, the results are coherent with Damasio's convincing theories about the existence of a background emotional system [4]. We demonstrated that phenomena such as body/world categorization and existence of a body map can evolve from a simple rule: self-survival. As a starting point to develop further Experiment 1, we suggest that the agent architecture should be updated. For instance we combined perception and action systems, allowing for affective responses to drive behavior, although we didn't incorporate at any level the fact that an affective response also has an internal feedback within the organism. A suggestion consists on creating a proprioceptive feedback system that can account for such interactions. At a different level we suggest the extension of the simulations scenario to a multi-agent system, allowing, for instance, for studies of emotion at different levels such as communicative or social. We also intend to apply our model to decision making tasks (e.g. music composition), as it allows to reduce the space state of choices, through an emotional categorization. Another interesting perspective comes from recent claims, specially Internal Robotics [6].

From a different perspective we are trying to understand how the different perceptual dimensions associated with music are perceived, integrated and organized, in such a way that they convey meanings as a whole. This task implies processes of segmentation and differentiation, as well as categorization. Moreover, in music, the temporal combination of the stimuli has a strong emotional effect. These aspects suggest that music can be a relevant source of information about the brain organization, accounting with new information for cognitive studies. If we can mathematically represent the basic aspects of musical emotions, we might also be able to bring that knowledge into computer models, using it also as a platform for emotion/cognition studies. Current work shows that musically induced arousal can be predicted by looking at the psychoacoustic properties of the stimuli. Future directions focus on formalizing the spatiotemporal dynamics of the stimuli, from their effect on arousal. We might be able to understand how these signals are organized and though obtain information of the nature of the stimuli that can trigger emotional responses.

ACKNOWLEDGMENTS

The authors acknowledge the financial support from the Portuguese Foundation for Science and Technology (FCT).

REFERENCES

[1] D. Canamero, "A hormonal model of emotions for behavior control", in *ECAL'97*, 1997.

[2] S. Gadanho and J. Hallam, "Robot learning driven by emotions", *Adaptive Behavior*, vol. 9, pp. 42–64, 2001.

[3] R. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state", *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1175–1191, 2001.

[4] A. Damasio, *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. Vintage, 2000

[5] E. Coutinho and A. Cangelosi, "The dynamics of music perception and emotional experience: a connectionist model", in *Proc. Int. Conf. on Music Perception and Cognition*, 2006.

[6] D. Parisi, "Internal robotics", *Connection Science*, vol. 16, no. 4, pp. 325–338, December 2004.

[7] J. Panksepp, "The neuro-evolutionary cusp between emotions and cognitions: Implications for understanding consciousness and the emergence of a unified mind science", *Consciousness & Emotion*, vol. 1, no. 1, pp. 15–54, 2000.

[8] K. Scherer, *Approaches to emotion*. Hillsdale: Erlbaum, 1984, ch. On the nature and function of emotion: a component process approach, pp. 293–317.

[9] P. Philippot, G. Chapelle, and S. Blairy, "Respiratory feedback in the generation of emotion", *Cognition and Emotion*, vol. 16, no. 5, pp. 605–627, 2002.

[10] S. Schachter, *The interaction of cognitive and physiological determinants of emotional state*, ser. Advances in experimental Social Psychology, L. Berkiwitz, Ed. New York: Academic Press, 1964, no. 1.

[11] B. Zillmann, *Transfer of Excitation in emotion behavior*, ser. Social psychophysiology, J. Cacioppo and R. Petty, Eds. Guilford, 1983.

[12] E. Coutinho, E. Miranda, A. Cangelosi, "*Towards a Model for Embodied Emotions*". In C. Bento, A. Cardoso & G. Dias (Eds.), *Proc. Portuguese Conf. on Artificial Intelligence*. Portugal, IEEE Press, pp. 54-63, 2005.

[13] E. T. Rolls, "Memory systems in the brain", *Annu. Rev. Psychol.*, vol. 51, pp. 599–630, 2000.

[14] D. S. Levine, "Seek Simplicity and Distrust it: Knowledge Maximization versus Effort
Minimization", International Conference on Integration of Knowledge Intensive Multi-Agent Systems (KIMAS'07), April 30 - May 3, 2007, Waltham, MA.

[15] L. I. Perlovsky, "Music - the First Principle", In J. Dimitrin, e-journal, *Musical Theatre*, http://www.ceo.spb.ru/libretto/kon_lan/ogl.shtml.

[16] A. Gabrielsson and E. Lindström, *The influence of musical structure on emotional expression*, In P. Juslin, and J. Sloboda (Eds.), Music and Emotion: Theory and Research (pp. 223-248). New York: Oxford University Press, 2001.

[17] R. J. Zatorre, "Music, the food of neuroscience?" *Nature*, vol. 434, pp. 312–315, 2005.

[18] S. Koelsch, T. Fritz, D. Cramon, K. Müller, and A. Friederici, "Towards a neural basis of music perception", *Trends in Cognitive Sciences*, vol. 9, pp. 578–584, 2006.

[19] H. G. Wieser, "Music and the brain: Lessons from brain diseases and some reflections on the "emotional" brain." *Annals. New York Academy of Science*, vol. 99, pp. 76–94, 2003.

[20] M. Clynes, *Sentics: the Touch of Emotions*. New York: Doubleday, 1978.

[21] P. Janata and S. T. Grafton, "Swinging in the brain: shared neural substrates for behaviors related to sequencing and music", *Nature Neuroscience*, vol. 6, pp. 682–687, 2003.

[22] A. J. Blood and R. J. Zatorre, "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion", *Proceedings of the National Academy of Sciences*, vol. 98, pp. 11818–11823, 2001.

[23] M. Korhonen, "Modeling continuous emotional appraisals of music using System Identification" [Online]. Available: http://www.sauna.org/kiulu/emotion.html.

[24] Schubert, E. (1999). *Measurement and time series analysis of emotion in music*. PhD thesis, Univ. of New South Wales.

Coutinho, E. & Cangelosi, A. (2006). The dynamics of music perception and emotional experience: a connectionist model. *In Baroni, R., Addessi, A. & Costa, M. (Eds.), Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC9).* Bologna, Italy: Bologna University Press.

9th International Conference on Music Perception and Cognition
Alma Mater Studiorum University of Bologna, August 22-26 2006

The dynamics of music perception and emotional experience: a connectionist model

Eduardo Coutinho
Adaptive Behaviour & Cognition Group
University of Plymouth
Plymouth, UK
eduardo.coutinho@plymouth.ac.uk

Angelo Cangelosi
Adaptive Behaviour & Cognition Group
University of Plymouth
Plymouth, UK

ABSTRACT

In this paper we present a methodological framework for the study of musical emotions, incorporating psychophysiological experiments and modelling techniques for data analysis. Our focus is restricted to the body implications as a possible source of information about the emotional experience, and responsible to certain levels of emotional engagement in music. We present and apply the use of spatiotemporal connectionist models, as a modelling technique. Simulation results using a simple recurrent network, demonstrate that our connectionist approach leads to a better fit of the simulated process, compared with previous models. We demonstrate that a spatiotemporal connectionist model trained on music and emotional rating data is capable of generalizing the level of arousal in response to novel music input. The model is also capable of identifying the main variables responsible for such an emotional rating behaviour.

Keywords
Music, emotion, brain, body, neural networks.

INTRODUCTION

As others (Panksepp & Bernatzky, 2002; Clynes, 1978) we believe that formal relationships between music psychoacoustic variables, body and brain dynamics exist. These embodiment factors are in part responsible for the emotional engagement in music. Without ignoring other processes involved in emotion induction through music (e.g. appraisal, memory), our focus is restricted to the body implications as a possible source of information about the emotional experience (e.g. proprioceptive feedback). Specifically we want to understand the relation between bodily arousal and emotional measurements, though requiring the use of both physiological and self-report frameworks.

In section 'Emotion' we discuss the physiological implications in the emotional experience, and we present our perspective. In the section 'Emotional Experience with Music', we refer to musical emotions and some of its prominent issues. Although the process that translates sounds into neural messages is nowadays quite well understood, several are the implications for the brain and the body during music listening that might be related with the emotion experience (Koelsch & Siebel, 2005). As in emotion experience, measurable effects are detectable on physiology (e.g. Iwanaga & Tsukamoto, 1997; Krumhansl, 1997), and brain dynamics (e.g. Patel & Balaban, 2000; Blood & Zatorre, 2001), during music listening. Clinical studies and music therapy explore the capacity of music to affect psychological and physiological states. They present evidence of helping patients with psycho-physiological symptoms, as in autism or Parkinson disease..

In: M. Baroni, A. R. Addessi, R. Caterina, M. Costa (2006) Proceedings of the 9th International Conference on Music Perception & Cognition (ICMPC9), Bologna/Italy, August 22-26 2006.©2006 The Society for Music Perception & Cognition (SMPC) and European Society for the Cognitive Sciences of Music (ESCOM). Copyright of the content of an individual paper is held by the primary (first-named) author of that paper. All rights reserved. No paper from this proceedings may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information retrieval systems, without permission in writing from the paper's primary author. No other part of this proceedings may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information retrieval system, without permission in writing from SMPC and ESCOM.

ISBN 88-7395-155-4 © 2006 ICMPC

We intend to model the temporal dynamics of music psychoacoustics, emotional ratings, and physiological states, in order to study the possible relations between them. Through an analysis of previous models, we will propose the use of spatiotemporal neural networks (Kremer, 2001), as a modelling technique. Their properties are discussed in the section 'Connectionist Models'. Finally, in section 'Case Study', we present a set of simulations and results using an Elman network (Elman, 1990), and we use these to discuss the theoretical framework proposed and the future modelling work.

EMOTION

Charles Darwin, William James, Walter Cannon, Wilhelm Wundt, Susanne Langer, among others, introduced great developments in the study of emotions, and are still considered as reference points for many emotions theorists. Although much discussion exists around the mechanisms of emotion, current research generally accepts that emotions can be described as a multi-modal mechanism, with several processes involved including appraisal, basic emotions, physiological responses, and subjective feeling states (Panksepp, 2001; Damasio, 2000; Scherer, 1984; Dolan, 2002).

It is important to stress that emotions differ in important aspects from other psychological processes. For instance, emotion is an embodied experience with specific behavioural patterns (facial expressions, autonomic arousal, etc.); it is less susceptible to our control and also expressed at the unconscious level (Damasio, 2000; Ekman, 1973); has the capacity to affect other cognitive processes (e.g. decision making), though not confined to the old sub-cortical structures in our brains (Damasio, 2000). For a review on these issues refer to Dolan (2002) and Panksepp (2001). For the interests of this article we focus on the implications of the body as a source of information or elicitation of the emotional experience.

Emotion as Arousal

The autonomic nervous system (ANS) regulates the body and its readiness for action. The correspondent physiological variations are referred as changes in arousal. The resultant patterns, as seen in the previous section, can elicit emotional states or experiences by interacting with the brain. Regarding specifically the role of physiological responses on the emotional experience and elicitation (and the interaction with the brain), there have been several models proposed. As discussed in Philippot, Chapelle, & Blairy (2002) paper on current research on the role of the body in the emotional experience, three main models of emotion can be distinguished. They identify (i) the "*undifferentiated arousal model*", (ii) the "*cognitive appraisal model*", and (iii) the "*central network model*".

The main idea behind the first model (Reisenzein, 1983; Schachter, 1964) is that body responses increase with emotional intensity, but their pattern is not differentiated across the different emotional states. One practical prediction of this model is that the perception of the emotional intensity can be influenced by the arousal intensity. The main finding of this research has been the fact that after the exposure to an arousing stimulus, the following emotional feeling state is intensified. This phenomenon is called "activation transfer" (Zillmann, 1983). The second model focuses on the body changes as a function of cognitive appraisal processes (Scherer, 1984), or action readiness (Frijda, 1986). In this line of research, the patterns of body changes are the combined result of the several cognitive appraisal components. The fact that the body itself might generate emotional states is quite marginalized in this model. This line of research has already been the one that we followed in our previous work (Coutinho, Miranda, & Cangelosi, 2005). It comprises several mechanisms of emotion induction, by considering all processes as part of a whole body/mind interactive system. Finally, from the third model perspective, emotions share different neural and cognitive mechanisms and pathways, and their pattern of interaction defines the emotional nature. In short, the patterns of body changes are differentiable across emotions. The activation of the body with a pattern related with a specific emotion will, in certain conditions, elicit that emotion (the peripheral feedback). This last process is automatic at an implicit level (Damasio, 1994).

In summary, the underlying idea common to all these mechanisms is that a specific emotion can be elicited by creating specific body state patterns (by manipulating the body), even outside the awareness of the individual. An event (appraised via cortical or subcortical routes) elicits physiological changes that facilitate action and expressive behaviour. These changes are accompanied by, and contribute to, an affective feeling state. Motoric and visceral feedback can contribute to the intensity and valence of an emotional experience: consciously or subconsciously, individuals use their body state as a clue as to the valence and intensity of the emotion they feel (Damasio, 1994; Damasio, 2000).

Peripheral feedback

In some very interesting experiments, Philippot et al. (2002) tried to find empirical evidence of emotion experience through physiological induction. To achieve such a task "physiological state should be manipulated, not in intensity but in quality, in order to observe the impact of such manipulations on the nature rather than the intensity of feeling states. Further, it should be established whether this effect occurs without individuals explicitly using body state as a source of information to determine their emotional feeling state" (Phillippot et al., 2002, p. 6). Respiration is the variable used, since it is considered appropriate for the study aims, as breathing is both under voluntary and automatic control. There is also empirical evidence pointing to the fact that respiratory patterns are associated with positive and negative feelings, and it is well known that respira-

ISBN 88-7395-155-4 © 2006 ICMPC

tion affects many other physiological responses like skin conductance or cardio-vascular changes, and that it is related with anxiety states.

The first experiment in Philippot et al. (2002) indicates that people experience respiratory changes that are subjectively differentiated across different types of emotions. Experiment 2 showed that differentiated emotional feeling states were induced by respiration manipulations without participants' awareness of the process. Philippot et al. claim that this is the first demonstration that the alteration of respiration is sufficient to induce emotion. Concluding, peripheral feedback can, at least in part, modulate the quality of emotional feelings, and this can be achieved without the person awareness of the process.

EMOTIONAL EXPERIENCE WITH MUSIC

Current research supports the hypothesis that music perception is a distributed process within the brain and is involved in an interactive spatiotemporal system of different neural networks (Zatorre, 2005; Koelsch, Fritz, Cramon, Müller, & Friederici, 2006), including even networks specialized for other processes (e.g. emotion, memory, language).

Music and brain dynamics

Experimental evidence suggests the participation of both hemispheres in music perception (Wieser, 2003), although it is possible to observe some interesting specializations. For instance the left hemisphere seems to be related to perception of timing and rhythm, while the right hemisphere specializes in pitch and timbre perception. Along with the spatiotemporal patterns of neural activity presented to the primary auditory system, the brain engages in other processes, for instance in the motor system (Thaut, Kenyon, Schauer, & McIntosh, 1999), language (Patel, 2003), emotion and reward related areas (Blood et al., 2001; Blood et al., 1999).

Some researchers (Panksepp, 2001; Clynes, 1978; Janata & Grafton, 2003) suggest that music derives its affective power from dynamic aspects of the brain systems. These usually control emotional processes and are distinct, but interacting, with cognitive processes. This hypothesis follows Susanne Langer's ideas about the existence of shared properties in patterns of physical or mental states, emotion, and music. Support for these ideas comes from research with brain damage patients. This shows that the emotional appreciation of music can be maintained even in the presence of severe perceptual and memorization deficits, though reinforcing the idea that sub-cortical mediation is involved in "emotional judgments" (Blood et al., 1999; 2001).

Due to these interactions certain basic mechanisms related to motivation/emotion in the brain can be elicited by music. This gives rise to the changes in the body and brain dynamics, and to the interference with ongoing mental and bodily processes (Panksepp et al., 2002; Patel et al., 2000). This multi-modal integration of musical and non-musical information might take place in the brain (Koelsh, 2005), opening a window for associations between the role of the body and the emotional experience during music listening. This indicates the existence of a possible relationship in the dynamics of musical emotion and the cognition of musical structure.

Music and body dynamics

Another quantifiable aspect of emotional responses to music is its effect on hormone levels in the body (Brownley, McMurray, & Hackney, 1995). There is evidence that music can lower levels of cortisol (associated with arousal and stress), and raise levels of melatonin (which can induce sleep). It can also cause the release of endorphins (Van der Ark & Ely, 1993), and can therefore help relieve pain. Krumhansl (1997), reports that sad excerpts are related to large changes in heart rate, blood pressure, skin conductance, and temperature, while fear excerpts are related with changes in the rate and amplitude of the blood flow. Happy ratings were associated with changes in respiration measurements. Although the correlation is fairly low and variable across individuals, some physiological changes related to musically induced emotions were found (Krumhansl, 2002).

In physiology, arousal is the term to define the body's readiness for action. Increased arousal is associated with increased heart rate, increased body temperature, increased respiration rate (increased oxygen consumption), and many other physiological changes. From a neurological perspective, increased arousal is associated with the release of adrenaline and noradrenaline, affecting for instance the amygdala. As a reference, the amygdala, deep in the limbic system, can influence cortical areas via feedback from proprioceptive, visceral or hormonal signals, via projections to various networks. In summary, hormones secreted in the body affect bodily processes (e.g. cardiovascular, muscular and immune systems) and the brain as well. Music can interact with both.

Peripheral feedback and Music

Nicola Dibben (2004) conducted a study to analyze the role of peripheral feedback in emotional experience with music. In a first experiment she tries to discover whether arousal causes intensification of emotional feeling when listening to music, comparing with the emotion thought to be express by the excerpt. Then she analyzes how the valence and arousal character of the music might mediate that process. Pulse rate was used as measure of arousal. To differentiate further variables, a second experiment was carried out to study the effect of the physiological arousal on emotion. It also looked at the origin of this effect, whether it was due to peripheral feedback or due to mood changes associated with the experiment. From the analysis of both experiments Dibben (2004) concludes that increased arousal influences listeners' experiences of emotion. Experiment 1 showed that increased physiological arousal intensified the domi-

ISBN 88-7395-155-4 © 2006 ICMPC

nant valence of emotions felt when listening to music, but not the arousal dimension of their emotional ratings. Experiment 2 indicated that if music has a positive valence, positive emotions feelings and thought are intensified more clearly than with negative dimension. Overall these results show that arousal intensifies the dominant valence response (in this case to music).

Measuring musical emotions

The findings presented above are particularly interesting as they demonstrate that the perception of the body state can be the source of experienced emotional feeling in music. From another perspective, the fact that the body state was manipulated shows that the source of the emotional arousal can be misattributed (in the above experiments they were attributed to music). This is consistent with the fact that listeners explore the environment to find clues for the type of emotion expressed by music, especially when it is not evident or complex (Juslin & Sloboda, 2001). Peripheral feedback may then be one of the sources used to classify that emotional experience.

Scherer (2004) posits some important questions on the study of musical emotions, mainly issue related to experimental frameworks. He suggests a combined methodology for the study of feelings induced by music integrating their cognitive and physiological effects. Frameworks to measure emotional experiences include discrete models (e.g. lists basic emotions: Ekman, 1999; Plutchik, 1991), eclectic approaches (e.g. Scherer, 2004), and dimensional models (e.g. Russell, 1989). In this article we focus on the last model, due to the temporal dynamics approach to emotion, as outlined earlier. For a discussion about the different methodologies, refer to Scherer (2004) and Schubert (1999).

The dimensional approach: arousal and valence

Wundt's (1897) initial dimensional model of emotion suggested the division of the emotion into three dimensions of feeling: pleasantness-unpleasantness, rest-activation, and tension-relaxation. This classification can be simplified using the two-dimensional space of arousal and valence. Both can be defined as subjective experiences (Russell, 1989). Arousal corresponds to a subjective state of feeling activated or deactivated (bodily activation); valence stands for a subjective feeling of pleasantness or unpleasantness (hedonic value). The conscious affective experience may be associated with a tendency to attend to the internal sensations associated with an affective experience, both of activation and of hedonic impact (Feldman, 1995).

Among the advantages of this model we can consider the simplicity (also for the experiment participants), and good reliability (Scherer, 2004). More recently Schubert (1999) has applied this concept to music creating the two dimensional EmotionSpace experimental software. While listening to music, participants were asked to continuously rate the emotion thought to be expressed by music. Each rating would correspond to a point in on the arousal/valence dimensional space.

Continuous measurements

Another important issue in music and emotion research regards the experimental methodology used to define the music stimulus. As pointed out by Schubert (2004), two main perspectives have been chosen for music stimulus in experiments: (i) the atomistic approach, based on the use of short auditory stimulus specifically produced for certain experiments, and (ii) the ecologically valid approach based on the use of "real" music. This idea of continuous measurements is consistent with the fact that music unfolds on time, since both body and brain engage in temporal processes by interacting with the musical cues. In line with Scherer (2004), this supports a framework to complement the study of the temporal dynamics of both physiological and brain processes, during music listening.. For example, rhythm and beat affect body rhythms and motor dynamics (e.g. Byers, 1976). A change in respiration rate due to musical rhythm, through the cardiovascular function, could affect several neurophysiologic systems (Boiten, Frijda, & Wientjes, 1994). This process is very similar to an emotion-induced arousal changes, in this case elicited by music.

We support the second approach based on ecological stimuli and continuous measurements. This is because continuous measurements of both music and emotion temporal dynamics better represent non linear relationships between body, music and emotions (Schubert, 1999). In the following part we will focus on a modelling techniques proposed to analyze the results from the framework suggested here.

SPATIOTEMPORAL MODELS

Several researchers (e.g. Gabrielsson & Lindström, 2001) agree on the point that the way musical elements are organized in time by the composer can evoke emotional responses in the listener. Parameters such as mood, physiological state, cultural background, and preferences of the listener influence this process. Other variables include performance style (Juslin, 2001) and musical style. This literature implies the existence of a causal, underlying relationship between musical features and emotional response. Hevner (1936) produced the first comprehensive work that attempted a systematic explanation of the relationship between musical features and perceived emotion (For a detailed review, see Gabrielsson & Lindström, 2001). This line of research found interesting relations between musical variables and elicited emotional states. However, the hypotheses proposed are not conclusive, especially due to the complexity of the psychoacoustic properties of music involved, and their linear and non-linear relations. We will now focus on the contribution of mathematical and computational modelling techniques for understanding the relationship between music elements and evoked emotions. In particular, we support the use of spatiotemporal models, i.e. approaches where the model at the same time includes a

ISBN 88-7395-155-4 © 2006 ICMPC

temporal dimension (e.g. musical sequences and continuous emotional ratings) and a spatial component (e.g. the parallel contribution of various music and psychoacoustic factors).

Previous models

Schubert (1999) and Korhonen (2004) were the first to study the interaction between music psychoacoustics and emotion ratings. They focus on the relationship between musical/psychoacoustic variables and emotional ratings using the continuous response methodology. Schubert proposes a methodology based on the use of combinations of time series analysis techniques (e.g. linear regression) to analyze the data and to model such process. Korhonen (2004) proposes the use of the System Identification technique (Ljung, 1999). Both authors focus on the issues of the complexity of the data and their possible nonlinear relationships. Schubert monitors melodic pitch, tempo, loudness, timbral sharpness, and texture. Then he applies first order autoregressive adjustments for serial correlation, building a regression model of emotional ratings and selected musical features. The extension of the model to a non-linear space with other possible sources of analysis is also suggested. As it will be discussed below, neural networks provide such an extension.

Korhonen (2004) extends the music features space and the musical repertoire. The experimental setup is similar to Schubert's one, and the modelling techniques consider the increased complexity of the data and generalization properties. The System Identification technique was chosen as it can model time-varying patterns, and it offers generalization properties. Again, for future work, the use of nonlinear models is suggested, with more generalization properties. Schubert highlights also the problem correlation between psychoacoustic variables, suggesting the identification of spatial patterns, along with temporal.

Spatiotemporal connectionist models

The study of the interaction between musical features and emotion ratings requires models that are capable of identifying spatiotemporal relationships among the data both in individual temporal patterns and in musical spatial (e.g. psychoacoustic) features correlations. This constitutes a complex dynamical system to which we can add generalization capabilities. Spatiotemporal connectionist models (Kremer, 2001) are an ideal methodology to investigate the dynamic spatiotemporal relationship in music and emotion..

Connectionist models (based on artificial neural networks) are computational paradigms inspired by the highly complex, nonlinear, and parallel information processing that occurs in the brain. A spatiotemporal connectionist networks can be defined as "a parallel distributed information processing structure that is capable of dealing with input data presented across time as well as space" (Kremer, 2001, pp. 2). There are several approaches to develop algorithms that achieve such a task, focusing on the dynamical processing of temporal patterns across a distributed system. Neural networks have been extensively used in computer music research (e.g. Todd & Loy, 1991; Martins & Miranda, 2006)

Neural Networks and Time: Elman networks

Neural networks that have to process time-based tasks, such as time series forecasting, often make use of recurrent connections to endow the network with a kind of dynamic memory. This way, the network can detect not only static patterns (not changing in time), but also add another dimension, containing temporal related information. Various proposals and architectures can be found in literature for time-based neural networks (see Kremer 2001 for a review). In this work we have selected Elman network (Elman, 1990), also called Simple Recurrent Network (SRN).

An Elman Neural Network (ENN) is based on the basic feed-forward architecture (multi-layer perceptron) with an additional layer called "context" or "memory" layer. The units in this layer receive a copy of the previous internal state of the hidden layer. They are connected back to the same hidden layer, through adjustable (learning) weights. These units endow the network with a dynamic memory, achieved through recursive access to past information. The basic functional assumption is that the next element in a time-series sequence can be predicted by accessing the previous hidden state of the system. This network has been extensively applied in areas such as language (e.g. Elman, 1990) and financial forecasting systems (e.g. Giles, Lawrence, & Tsoi, 2001), among others.

We have chosen an ENN because it allows the simulation of a task in which the input time series consists of a music piece, and the output a rating of emotional state (e.g. arousal). Moreover, Elman networks have very good generalization capabilities.

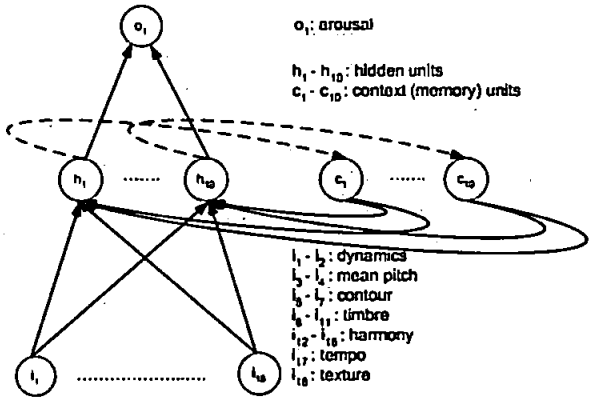


Figure 1. Neural network architecture (ENN).

CASE STUDY

Our work consists on the simulation modelling of Korhonen's (2004) experimental data. He used 6 pieces of classi-

ISBN 88-7395-155-4 © 2006 ICMPC

cal music[1] for his experiments. Volunteers used a continuous arousal and valence scales (with a sample rate of 1s), to rate the emotion thought to be expressed by the music. We use the same 18 psychoacoustic variables chosen by Korhonen to encode the musical input to the system, and, for the purpose of this preliminary analysis of the model, as output, we selected only the arousal ratings (the focus of this preliminary analysis of the model). Music 1 to 5 are used to train the system, while Music 6 is used to test the generalization response to novel stimuli.

The music input stimuli were encoded into 18 psychoacoustic variables, clustered into 7 major groups: dynamics (loudness), mean pitch (centroid), pitch variation (contour), timbre, harmony (western consonance/dissonance), tempo, and texture (no. instruments playing). These correspond to the inputs of the ENN, while the single output refers to arousal. This way we are trying to find the relations between music and emotion appraisal, assuming the last one as a non-linear function of the former. The network architecture and its specifications are shown in Fig. 1.

Methodology and Simulations

We carried out two sets of simulations to evaluate the connectionist framework proposed here, and to be able to compare it with the other referenced models. Experiment 1 focuses on a quantitative analysis of the model performance in learning (training data) and generalizing (validation data) the musical temporal sequences, and relating them with emotional appraisals. Experiment 2 consists in evaluating the relative contribution of each individual group of psychoacoustic music variables to arousal dynamics, in order to find relationships between them. For all experiments we present the results of 5 representative replications, each corresponding to the training of an ENN with different randomly initialized weights. In Experiment 1 we train we trained the ENN with Music 1 to 5, and we calculate the errors for the train data and for the novel data of Music 6 (see Table 1 and Fig. 2). Then, in Experiment 2, we removed the input values of each individual psychoacoustic group from the network, in a set of 7 simulations (see Table 1 and Fig. 2). This allowed us to analyze the influence of each individual psychoacoustic lag on the output prediction, therefore identifying their contribution to the arousal rating. All simulations were run for 2000 epochs (803 training sweeps per epoch, i.e. the combined total length of the 5 training music pieces), with a learning rate of 0.01 and momentum of 0. The context layer was reset every 20 training sweeps (corresponding to 20s of music).

[1] Music 1 - Concierto de Aranjuez – Adagio (Rodrigo)
Music 2 - Fanfare for the Common Man (Copland)
Music 3 - Moonlight Sonata – Adagio Sostenuto (Beethoven)
Music 4 - Peer Gynt – Morning (Grieg)
Music 5 - Pizzicato Polka (Strauss)
Music 6 - Piano Concerto No. 1 – Allegro Maestoso (Liszt)

Results

The network learning errors are calculated using the Root Mean Square (RMS) error. This measure gives us an indication on how well the network is capable of representing the desired output. We then can calculate the squared multiple correlation coefficient R^2 (or fit), as in Eq. 1, where N corresponds to the length of the time series, and $y(t)$ to the ideal arousal value at the output (the train target at time t). This measure allows us to discuss the amount of output variation that it is explained by the model, and to compare it with the performance of Schubert's and Korhonen's models. The average training and generalization errors of experiment 1 are presented in the first column (*All variables*) of Fig. 2. The individual errors for the simulations of Experiment 2 correspond to the other columns in the histogram of Fig. 2.

From Table 1, we can see the significantly improved performance of the ENN model (R^2 values of 97% and 98%) compared with the other two approaches (R^2 values below 90%). This supports the validity of the spatiotemporal connectionist model. Another important result is the capability of the network to generalize to new novel music input, as it is the case with Music 6. Similar high performance to that of the training was obtained for the validation tests. This indicates that the neural network was able to infer the underlying dynamics for rating the arousal level produced by the music input.

(Equation 1)

$$R^2 = \left(1 - \frac{RMS^2}{\frac{1}{N}\sum_{t=1}^{N}|y(t)|^2}\right) * 100\%$$

Table 1. Experiment 1 (all variables in input): Training error and correlation coefficient in the ENN and Korhonen's (2004) and Schubert's (1999) models.

Music	Average RMS	R^2 Exp. 1	R^2 Korhonen	R^2 Schubert
Train data				
1	0.069	98	90	57
2	0.075	97	-170	n. a.
3	0.058	98	25	n. a.
4	0.097	96	66	67
5	0.060	98	65	36
6	n. a.	n. a.	86.5	n. a.
Novel data				
6	0.092	97	n. a.	n. a.

ISBN 88-7395-155-4 © 2006 ICMPC

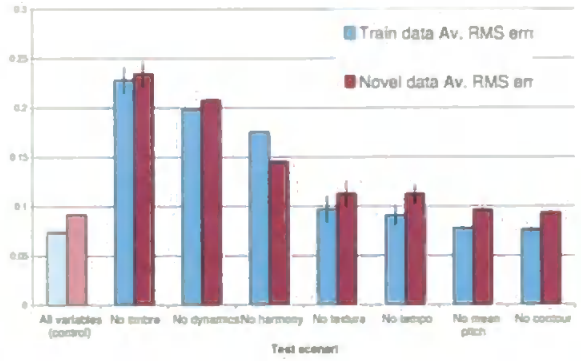


Figure 2. Experiment 2: train and test average RMS and standard errors.

Another analysis is reported in Fig. 2. As presented before we wanted to understand to which extent specific musical variables might be closely related with emotional judgments. For this we conducted individual analysis for each musical variable group. The histogram shows 3 cases of relevant error increase when removing these specific psychoacoustic groups: dynamics (loudness), timbre, and harmony. The significant drop in performance (i.e. error increase) indicates the stronger correlation between these variables and arousal ratings, suggesting their close relatedness for the repertoire and population used. This result agrees with research in music and emotion, as extensively reviewed by Schubert (1999). In his work he reports the strong relationship between arousal and loudness, as well with tempo. Here we report that timbre as well as loudness and harmony, can explain great part of the variations in arousal ratings. Texture and tempo didn't have a relevant impact on the process; also mean pitch and melody contour do not seem to affect the model performance. Nevertheless our results support and extend Schubert thesis. We will apply more extensive methodologies to analyze their temporal dynamics, variables correlations, and their relation with musical context.

SUMMARY AND DEVELOPMENTS

In this paper we presented a methodological framework for the study of musical emotions, incorporating psychophysiological experiments and modelling techniques for data analysis. Then our focus is restricted to the body implications as a possible source of information about the emotional experience, and responsible to certain levels of emotional engagement in music. Along with this, we presented a framework to model the temporal dynamics of music psychoacoustics, emotional ratings, and physiological states, in order to study the possible relations between them, through an analysis of previous models. We present and apply the use of spatiotemporal connectionist models, as a modelling technique.

Preliminary results using a simple recurrent network, improve previous models results. Moreover they open an interesting window for temporal dynamical analysis of a nonlinear system, as it seems to be the case of our investigation. In summary we demonstrated that a spatiotemporal connectionist model trained on music and emotional rating data is capable of generalizing the level of arousal in response to novel music input. The model is also capable of identifying the main variables responsible for such an emotional rating behaviour.

The next steps include a complete analysis of the data and improvement of the model. Along with that we are working on the psycho-physiological framework to establish a set of experiments, exploring also new techniques. We intend to explore the framework proposed in order to obtain better experimental setups. It is interesting to analyze to what extent this can be also applied to a more diverse musical scenario. One of the difficulties relies on colinearity of musical stimuli, affecting the model capability to represent the desired process. Namely the case of tempo is one of the evident facts in this article, since apparently no relevant quantitative correlations were found with emotional ratings. We believe that this can be related with both with representation or variables co-linearity issues. Nevertheless an extensive analysis is required to find if effectively this and other variables are correlated with the emotional reports in this experiment. Another important issue is the music representation, since the transformation for the neural network inputs may also introduce some problems. Current developments include strategies to solve both the problem of music stimuli and their diversity of contents and contexts (allowing individual and mutual correlation analysis), and the music representation. These issues are extremely relevant to endow our framework with a reliable basis to conclude about the influence of music on emotional ratings, in a more "universal" basis. Finally, valence will also be modelled, although some preliminary analysis has shown this to be a more difficult task to achieve. We believe that physiological measurements will highlight and facilitate this process, helping to detect and ground important effects. Another possibility after the complete analysis is to use different dimensions for the emotional ratings.

AKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of the Portuguese Foundation for Science and Technology (FCT, Portugal). The acknowledgements are extended to Mark Korhonen for making available his experimental data, and also to Guido Bugmann, Andrew Hennell and Roman Borisyuk for their relevant comments and helpful suggestions.

REFERENCES

Bateson & P. Kopfer (Eds.). *Perspectives in Ethology* (pp. 135-164). New York: Plenum.

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

Blood, A. J. & 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, 98(20)*, 11818-11823.

Boiten, F. A., Frijda, N. H. & Wientjes, C. J. (1994). Emotions and Respiratory Patterns: review and critical analysis. *Int. J. Psychophysiology, 17*, 103-128.

Brownley K.A., McMurray ,R.G. & Hackney, A.C. (1995). Effects of music on physiological and affective responses to graded treadmill exercise in trained and untrained runners. *Int. Journal of Psychophysiology, 19(3)*,193-201

Byers, P. (1976). Biological rhythms as information channels in interpersonal communication behavior. In P.

Clynes, M (1978). *Sentics: the Touch of Emotions*. New York: Doubleday.

Coutinho, E., Miranda, E. & Cangelosi, A. (2005). Towards a Model for Embodied Emotions. *Proceedings of the Workshop on Affective Computing: Towards Affective Intelligent Systems (AC 2005)*, Covilhã (Portugal).

Damasio, A. (1994). *Descarte's error: emotion , reason and the human brain*. New York: Grosset/Putnam Books.

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.

Dolan, R. J. (2002). Emotion, Cognition, and Behavior. *Science 298(5596)*, 1191-1194.

Ekman, P. (1973). *Darwin and facial expression: A century of research in review*. New York: Academic Press.

Ekman, P. (1999). Basic Emotions. In T. Dalgleish & T. Power (Eds.). *The Handbook of Cognition and Emotion*. Sussex: John Wiley and Sons, Ltd.

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

Feldman, L. A. (1995). Valence-focus and arousal-focus: Individual differences in the structure of affective experience. *Journal of Personality and Social Psychology, 69*, 153-166.

Frijda, N. H. (1986). *The Emotions*. Cambridge: Cambridge University Press.

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In P. N. Juslin, & J. A. Sloboda (Eds.). *Music and emotion: Theory and Research* (pp. 223-248). New York: Oxford University Press.

Giles, C. L., Lawrence, S. & Tsoi, A. C. (2001). Noisy Time Series Prediction Using a Recurrent Neural Network and Grammatical Inference. *Machine Learning, 44(1/2)*, 161-183.

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

Ljung, L. (1999). *System Identification: theory for the user*. New Jersey: Prentice-Hall.

Iwanaga, M. & Tsukamoto, M. (1997) Effects of excitative and sedative music on subjective and physiological relaxation. Percept Mot Skills; 85: 287-296.

Janata, P. & Grafton, S. T. (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nature Neuroscience, 6(7)*, 682-687.

Juslin, P. N. & Sloboda, J. A. (2001). Psychological perspectives on music and emotion. In P. N. Juslin & J. A. Sloboda (Eds.). *Music and emotion: theory and research* (pp. 71-104). New York: Oxford University Press.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin, & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 305-333). New York: Oxford University Press.

Koelsch, S. & Siebel, W.A (2005).Towards a neural basis of music perception. *Trends in Cognitive Sciences, 9(12)*, 578-584.

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

Kremer S. C. (2001). Spatiotemporal Connectionist Networks: A Taxonomy and Review. *Neural Computation, 13*, 249-306.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology, 51(4)*, 336-353.

Krumhansl, C. L. (2002). Music: a link between cognition and emotion. *Current Directions in Psychological Science, 11*, 45-50.

Korhonen, M. (2004). *Modeling Continuous Emotional Appraisals of Music Using System Identification*. Phd thesis, University of Waterloo.

Korhonen, M. (2004). *Modeling Continuous Emotional Appraisals of Music Using System Identification*. Retrieved August 3, 2005, from http://www.sauna.org/kiulu/emotion.html.

Martins, J. M. and Miranda, E. R. (2006). A Connectionist Architecture for the Evolution of Rhythms. *Proceedings of EvoWorkshops 2006, Lecture Notes in Computer Sciente*. Berlin: Springer-Verlag.

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

Panksepp, J. (2001). The neuro-evolutionary cusp between emotions and cognitions. *Evolution and cognition, 7(2)*, 141-163.

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

Patel, A.D. & Balaban, E. (2000). Temporal patterns of human cortical activity reflect tone sequence structure. *Nature, 404*, 80-84.

Patel, A. (2003). Language, Music, syntax and the brain. *Nature Neuroscience, 6(7)*, 674-681.

Philippot, P., Chapelle, G. & Blairy, S. (2002). Respiratory feedback in the generation of emotion. *Cognition and Emotion 16(5)*, 605-627.

Plutchik, R. (1991). *The Emotions*. Lanham, MD: University Press of America.

Reisenzein, R. (1983). The Schachter theory of emotion: two decades later. *Psychological bulletin, 94*, 239-264.

Russell, J. A. (1989). Measures of emotion. In R. Plutchik & H. Kellerman (Eds.). *Emotion:Theory, research, and experience* (4, pp. 83-111). Toronto: Academic.

Schachter, S. (1964). The interaction of cognitive and psysiological determinants of emotional state. In L. Berkiwitz (Ed.). *Advances in experimental Social Psychology* (1, 44-79). New York: Academic Press.

Scherer, K. R. (1984). On the nature and function of emotion: a component process approach. In K. Scherer and P. Ekman (ed.). *Approaches to emotion* (pp. 293-317). Hillsdale: Erlbaum.

ISBN 88-7395-155-4 © 2006 ICMPC

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research, 33(3)*, 239-251.

Schubert, E. (1999). *Measurement and time series analysis of emotion in music*. PhD thesis, University of New Southy Wales.

Thaut, M. H., Kenyon, G. P., Schauer, M. L. & McIntosh G. C. (1999). The connection between rhythmicity and brain function. *IEEE Engineering in Medicine and Biology, 18(2)*. 101-108.

Todd, P. M. & Loy, D. G. (Eds.) (1991). *Music and connectionism*. Cambridge, MA: MIT Press.

Van der Ark, S. D. & Ely, D. (1993). Cortisol, biochemical, and galvanic skin responses to music stimuli of different preference values by college students in biology and music. *Perceptual and Motor Skills, 77(1)*, 227-234.

Wieser, H. G. (2003). Music and the Brain: Lessons from Brain Diseases and Some Reflections on the "Emotional" Brain. *Annals. N.Y. Academy of Science*, 99, 76–94.

Wundt, W. (1897). *Outlines of Psychology*. Leipzig: Wilhelm Engelmann.

Zatorre, R. J. (2005). Music, the food of neuroscience?. *Nature, 434*, 312-315.

Zillmann, D. (1983). Transfer of Excitation in emotion behavior. In J.T. Cacioppo & R.E. Petty (Ed.), *Social psychophysiology* (215-240). New York: Guilford.

ISBN 88-7395-155-4 © 2006 ICMPC

Coutinho, E., Miranda, E. & Silva, P. (2005). Evolving emotional behaviour for expressive performance of music. *In Panayiotopoulos, T., Gratch, J., Aylett, R., Ballin, D., Olivier, P. & Rist, T. (Eds.), Intelligent Virtual Agents: Proceedings of the 5th International Working Conference (IVA2005)* . Berlin (Germany): Springer-Verlag (LNAI 3661), pp. 147-147.

Evolving Emotional Behaviour for Expressive Performance of Music

Eduardo Coutinho, Eduardo Reck Miranda, and Patricio da Silva

Future Music Lab - University of Plymouth,
206 Smeaton Building – Drake Circus – Plymouth PL4 8AA – United Kingdom
{eduardo.coutinho, eduardo.miranda,
patricio.dasilva} @plymouth.ac.uk

Abstract. Today computers can be programmed to compose music automatically, using techniques ranging from rule-based to evolutionary computation (e.g., genetic algorithms and cellular automata). However, we lack good techniques for programming the computer to play or interpret music with expression. Expression in music is largely associated with emotions. Therefore we are looking into the possibility of programming computer music systems with emotions. We are addressing this problem from an A-Life perspective combined with recent discoveries in the neurosciences with respect to emotion.

Antonio Damasio refers to the importance of emotions to assist an individual to maintain survival, as they seem to be an important mechanism for adaptation and decision-making. Specifically, environmental events of value should be susceptible to preferential perceptual processing, regarding their pleasant or unpleasant. This approach assumes the existence of neural pathways that facilitate survival. Stable emotional systems should then emerge from self-regulatory homeostatic processes.

We implemented a system consisting of an agent that inhabits an environment containing with a number of different objects. These objects cause different physiological reactions to the agent. The internal body state of the agent is defined by a set of internal drives and a set of physiological variables that vary as the agent interacts with the objects it encounters in the environment. The agent is controlled by a feed-forward neural network that integrates visual input with information about its internal states. The network learns through a reinforcement-learning algorithm, derivate from different body states, due to pleasant or unpleasant stimuli.

The playback of musical recordings in MIDI format is steered by the physiological variables of the agent in different phases of the adaptation process.

The behaviour of the system is coherent with Damasio's theory of background emotional system. It demonstrates that specific phenomena, such as body/world categorization and existence of a body map, can evolve from a simple rule: self-survival in the environment. Currently, we are in the process of defining a system of higher-level emotional states (or foreground system) that will operate in social contexts; i.e., with several agents in the environment reacting to objects and interacting with each other.

T. Panayiotopoulos et al. (Eds.): IVA 2005, LNCS 3661, p. 497, 2005.
© Springer-Verlag Berlin Heidelberg 2005

Coutinho, E., Miranda, E. & Cangelosi, A. (2005). Towards a Model for Embodied Emotions. *In Bento, C., Cardoso, A. & Dias, G. (Eds.), Proceedings of the Portuguese Conference on Artificial Intelligence (EPIA2005).* Covilhã (Portugal): IEEE Press, pp. 54-63.

Towards a Model for Embodied Emotions

Eduardo Coutinho*†, Eduardo R. Miranda* and Angelo Cangelosi†
*Future Music Lab
†Adaptive Behaviour and Cognition Research Group
School of Computing, Communications & Electronics
University of Plymouth
Drake Circus
Plymouth PL4 8AA
Email: {eduardo.coutinho.eduardo.miranda,angelo.cangelosi}@plymouth.ac.uk

Abstract— **We are interested in developing A-Life-like models to study the evolution of emotional systems in artificial worlds inhabited by autonomous agents. This paper focuses on the emotional component of an agent at its very basic physical level. We adopt an evolutionary perspective by modelling the agent based on biologically plausible principles, whereby Emotions emerge from homeostatic mechanisms. We suggest that the agent should be embodied so as to allow its behaviour to be affected by low-level physical tasks. By embodiment we mean that the agent has a virtual physical body whose states can be sensed by the agent itself. The simulations show the emergence of a stable emotional system with emotional contexts resulting from dynamical categorization of objects in the world. This proved to be effective and versatile enough to allow the agent to adapt itself to unknown world configurations. The results are coherent with Antonio Damasio's theory of background emotional system [1]. We demonstrate that body/world categorizations and body maps can evolve from a simple rule: self-survival.**

I. Introduction

The importance of Emotions has been emphasized throughout the years in several areas of research. An interesting path has been traced by several researchers. There are findings in neuroscience, psychology and cognitive sciences indicating the surprising role of Emotions in intelligent behaviour. Specially interesting for us are the studies looking at physiological interferences, and the relation between body and affective states, as well to the evolutionary mechanisms. Emotions have an important role in behaviour and adaptation in biological systems.

In our modeling approach we share the neurobiological and evolutionary perspectives to Emotions [1], as discussed in the following sections.

A. Our perspective on Emotions

Going back to the 19^{th} century we find the earliest scientific studies on Emotions: Charles Darwin'[2] observations about bodily expression of Emotions, William James' [3] search for the meaning of emotion and Wilhelm Wundt's [4] appeal for the importance of including Emotions among the research topics in psychology studies. But for many years, studies on behaviour focused on higher level processes of mind, discarding Emotions altogether [5]. Still the ideas changed considerably throughout time, and nowadays Emotions are the focus of many researchers.

The line connecting mind and body, and the role played by Emotions in rationality came emphasized after Walter Cannon [6]. He suggested that there are neural paths from our senses that flow in two directions - the experience of an emotion and the physiological responses occur together. Later Silvan Tomkins [7], [8], Robert Plutchik [9], [10] and Carroll Izard [11], [12], [13] developed similar evolutionary theories of Emotions. They claimed that Emotions are a group of identical processes of certain brain structures and that each of them has a unique concrete emotional content, reinforcing their importance. Paul Ekman proposed the basic (and universal) Emotions [14], based on cross-cultural studies [15]. These ideas were widely accepted in evolutionary, behavioural and cross-cultural studies, by their proven ability to facilitate adaptive responses.

Important insights come from Antonio Damasio [16][1][17], who brought to the discussion some strong neurobiological evidence, mainly exploring the connectivity between body and mind. He suggests that, the processes of Emotion and Feeling are part of the neural machinery for biological regulation, whose core is formed by homeostatic controls, drives and instincts. Survival mechanisms are related this way to Emotions and feelings, in the sense that they are regulated by the same mechanisms. Emotions are complicated collections of chemical and neural responses, forming a pattern; all Emotions have some regulatory role to play, leading in one way or another to the creation of circumstances advantageous to the organism exhibiting the phenomenon. The biological function of Emotions can be divided in two: the production of a specific reaction to the inducing situation (e.g. run away in the presence of danger), and the regulation of the internal state of the organism such that it can be prepared for the specific reaction (e.g. increased blood flow to the arteries in the legs so that muscles receive extra oxygen and glucose, in order to escape faster). Emotions are inseparable from the idea of reward or punishment, of pleasure or pain, of approach or withdrawal, or personal advantage or disadvantage.

From a neurobiological perspective, the sequence of events in the process of Emotion, can be summarized as:

1) engagement of the organism by an inducer of emotion;
2) signals consequent to the processing of the object's image activate all the neural sites that are prepared to

respond to the particular class of inducer to which the object belongs. These sites have been preset innately, although past experience has modulated the manner in which they are likely to respond;

3) emotion induction sites trigger a number of signals toward other brain sites (for instance, monoamine nuclei, somatosensory cortices, cingulated cortices) and toward the body (for instance, viscera, endocrine glands).

Edmund Rolls [18], as Antonio Damasio [1], underlines the division of two concepts: Emotion and Feeling. The first corresponds to states derived from reinforcement stimuli; the second represents the real "feeling". A reinforcement signal brings information about reward and punishment. By the internal representation of an object, already biologically qualified as an emotion inducer (or acquired), generalization and association processes occur (e.g. fear: one snake ... all snakes). A perception followed by an emotional reaction can be the activation of a representation. The brain is modulated by reward/punishment processes, and its goal is to maximize rewards and minimize punishments [18].

We share some of Rolls ideas regarding the reward/punishment system, from the perspective in which the evolution of higher brain systems were guided by previous neurobiological predispositions [19]. Other implications of his ideas, specially in the relation between Emotion and Memory, are not part of our discussion. Nevertheless, this gives rise to an interesting discussion about interaction between affect and logic, Emotions and cognition. Sustained for an evolutionary perspective certain organizational principles in the brain might reflect emotional states.

For further details and references on Emotions, cognition and behaviour, please refer to [1] [19] and [20].

B. Arousal and Valence

An interesting approach to Emotions is the Dimensional Approach. In short, an emotion has at least two qualities: valence (pleasantness or hedonic value) and arousal (bodily activation). Both may be defined as subjective experiences [21]. Valence is a subjective feeling of pleasantness or unpleasantness; arousal is a subjective state of feeling activated or deactivated. A two dimensional model has been proposed to reflect the degree to which different individuals incorporate subjective experiences of valence and arousal into their emotional experiences [22]. In Fig. 1, there is a possible representation of this two-dimensional space.

C. Concept and Hypothesis

Due to a progressive change in theoretical studies in a broad range of areas, models of cognition, attention and behaviour now frequently include Emotions as part of the behavioural system. Emotional cues as states that might influence behaviour and adaptation is an idea that became stronger in the last decades, and gained special attention in computational models of cognition and behaviour [23], [24], [25], [26], [27], to cite but a few. Different perspectives were adopted when working with Emotions. While some of these models are

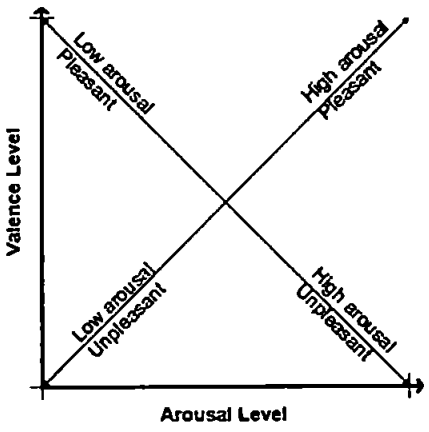


Fig. 1. Arousal-Valence Space.

inspired by different properties of an emotional system for task solving issues, and specific application (e.g. using facial expressions for social engagement), we are interested in using computational models to understand the basic mechanisms of the emotional systems. We are interested in studying how these systems evolved, which mechanisms do they use, and the role of the body.

We created a conceptual A-Life model to implement artificial worlds inhabited by autonomous emotional agents, modelling the agent based on biologically plausible principles. We focused on the idea of having an embodiment (in the sense that the agent has a virtual physical body whose states can be sensed by the agent itself) so that low level tasks (e.g. satiate body needs) influence its overall performance, by affecting its behaviour. A neural network endows the agent with cognitive capabilities, processing information related with its body, and its environment. The agent's emotional state is mirrored into a set Background Emotions. This term is used by Damasio [1] for the responses caused by "...certain conditions of internal state engendered by ongoing physiological processes or by the organism's interactions with the environment or both"[1].

The agent learns through a reward and punishment system, to adapt itself to the environment by interacting with it. Our algorithm is inspired on Rolls' "Stimulus-Reinforcement Association Learning" [28]: (...) stimuli or events which, if their occurrence, termination, or omission is made contingent upon the making of a response, alter the probability of the future emission of that response. Moreover, some stimuli are unlearned or primary reinforcers (e.g. pain), while others may become reinforced by learning, because of their association with such primary reinforcers: the secondary reinforcers"[2]. We use the arousal/valence dimensions referred in the previous subsection, to scale and quantify the interactions results. This way we intend to create the basis of we believe to be essential for the emergence of an emotional system (in the sense

[1] page 52.
[2] page 601.

that acts such like mechanism): a body, a brain, and their interaction.

In this model, Emotions act as an adaptation mechanism. We implemented the agents' background structure, which will eventually allow us to contextualize the foreground emotional system (and the so called Basic Emotions [14]), in order to define the agents's foreground (to use Damasio' terminology) emotional state.

The aim of this exercise is to design an embodied agent-based cognitive model and establish how an emotional system can emerge from self-regulatory Homeostatic Processes, by the interaction between a body and a brain. For that we propose a model of these biological mechanisms. The objective is to understand the role and the importance of Emotions in self-survival tasks; hence one of the reasons to implement a single-agent system at this stage. We also are interested in studying how the regulation of the Homeostatic Processes can influence world categorization and decision making (currently at a low level and for single tasks). We also analyze how Emotions act as a system of internal rewards, that preserve the system, and permit continuous adaptation process in self-survival tasks, by signalling and scaling pleasant or unpleasant interactions.

II. The Simulation Environment

The simulation environment (programmed in C++) is represented as a two-dimensional world populated by several objects (See Fig. 2). An autonomous agent inhabits this artificial world and is able to move within the borders that define its limits. The objects have different representations (see Table I). Each one of them is related with a physiological interference. We also created obstacles in the world that, in the simplest case, are only the topological limits (or borders) of the world: these borders are considered as sources of pain.

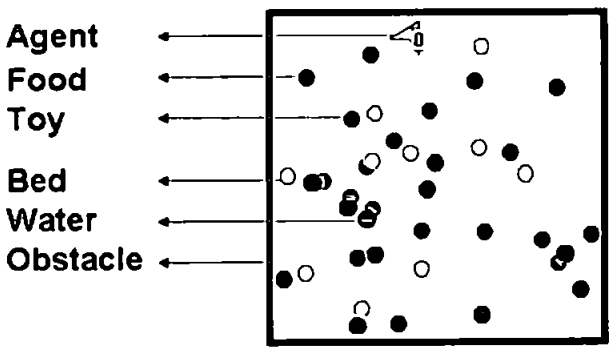


Fig. 2. The artificial world.

In short, we have an autonomous agent, which has to adapt itself to a world by controlling self-survival tasks, and by attributing emotional meanings to objects; objects might have different meanings for different situations. No representation or meaning of these objects is given to the agent beforehand. In order to design the autonomous agent, its structure includes cognitive, emotional and embodiment systems, endowing the agent with capabilities of interacting with the world, sense the body and the world, and learn from that.

Object color	Representation	Physiological interference	Motivation
Red	Food	Increase Blood Sugar	Eat
Green	Bed	Increase Energy	Rest
Blue	Obstacles	Increase Pain	-
White	Water	Increase Vascular Volume	Drink
Pink	Toy	Increase Endorphine	Play

TABLE I
OBJECTS REPRESENTATION AND INTERFERENCES.

III. The Autonomous Agent

We developed a versatile structure for the agent, which will allow us to extend it as the project develops; e.g., the creation of a truly interactive artificial society. Currently, the agent comprises three main formalized systems: Perceptual System, Nervous System and Motor System. These structures will be developed at a later stage in order to incorporate more complex interaction interfaces between agents and interaction with more complex objects (possibly associated with higher level tasks in the world). As a basis to create this system, information is stored as "genomic" structure, that represents the characteristics of the body and nervous system of the agent. We emphasize the fact that no structure was explicit coded for Emotions: we are studying the possibility of an emergent phenomena due to the modelled biological system.

A. Perceptual System

In order to perceive the world, an agent contains a retina (here represented as a color array) that resembles a biological retina on a functional level, inspired in LIVIA [29] and GAIA [30]. It senses a bitmap world (environment) through a *ray tracing* algorithm, which is inspired on the photons travel from the light-emitting objects to the retina. At each time step, an agent fetches in a certain number of directions for visual input from its world. Each light ray that hits the sensing cells is traced to its origin in order to determine its intensity and colour, which feeds directly a color array. This array is then relayed to the nervous system. Each of the five colors is mapped into 4 neurons in the neural network input, as seen in Fig. 3. The directions in which the agent fetches the objects is calculated by Equation 1 (where i stands for the direction index, and α_i the fetching angle for each direction).

$$\alpha_i = \frac{2 \cdot \pi \cdot i}{retina_size} \quad (1)$$

Another characteristic is the attenuation of visual cues in function of the distance of the objects in relation to the agent. There is a maximum value for distances in which the agent can see (*sight_range*). Colors are linearly attenuated according to their distance from the agent. That is, a *foggy world* was created. Currently, the agent has a sight range of 80 pixels in a 250x250 bi-dimensional world. Both *retina_size* (number of fetching angles) and *sight_range* are coded into the "genomic" structure of the agent.

B. Nervous System

The nervous system includes a feed-forward neural network (NN) with a genetically encoded structure (fixed during life-time). The neural network is organized in layers: an input layer (two groups: retina and body sensors), an output layer (two groups: motivations and motor control) and a hidden layer (with excitatory-only and inhibitory-only neurons). We distinguish between excitatory and inhibitory hidden groups due to the fact that the agent will have to perform tasks related to the activation or inhibition of certain behaviours. In the current version of the nervous system, each neuron of a layer connects to all neurons of the following layer. The inputs are propagated to the output through the synapses, processing one layer at the time from the input to the output. Each area has projections (a group of synapses) to any other area of the following layer. In Fig. 3, we present the NN architecture. Table II shows the current values that define the NN architecture.

Group	Number Neurons	Number of Synapses
Retina	20	20 · 16 = 320
Body Sensors (Drives)	5	5 · 16 = 80
Hidden Layer	16 (8 excit.+8 inhib.)	16 · 9 = 144
Motivations	6	0
Motor Control	3	0

TABLE II
THE VALUES FOR THE NEURAL NETWORK ARCHITECTURE.

The activation function for the input neurons is presented in Equation 2, while the activation for all other neurons is calculated by Equation 3.

$$F_{input} = \left(\frac{\tan^{-1}(x)}{\pi/2} \right) \tag{2}$$

$$F = \left(\frac{1}{1 + \exp^{-a \cdot x}} \right) \tag{3}$$

C. Motor System

The concept of Emotions, and body relatedness imply the notion of an interactive embodiment system.

We created a simple structure for this: the agent controls a motor system through linear and angular speed signals, allowing it to travel around the world (including obstacle avoidance and object interaction). These signals are provided by the neural network, which means that motor skills also have to be learned. With this capabilities the agent will be able to navigate in its environment, approaching or avoiding certain states. This constitutes an important aspect for this study.

D. Embodied Emotional Process

As a starting point, we shall highlight one important aspect: rather than model emotional systems, we are interested in modeling the basic biological interactive elements (body and brain) from where Emotions and Feelings might emerge. In other words, we are interested in modelling the conditions for the emergence of Emotions, instead of programming Emotions.

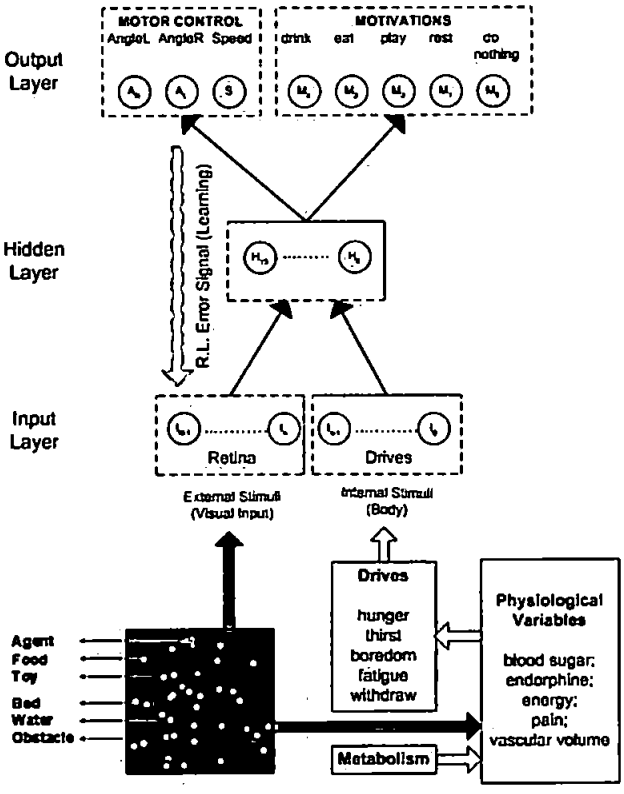


Fig. 3. Body/Brain interaction. System overview.

Fig. 3, gives a simple overview of the system. Inspired by Izard's work [13], we categorized the stimuli sensed by the agent as follows:

1) Somatic - body state (physiological data, drives);
2) World Perception (vision);
3) Body - external interactions (pain).

The agent perceives the world through a retina, and this signal is used to feed a set of input layers of the NN, together with its internal body state.

The body state consists of a map of the agent's body. We introduce a set of Physiological Variables into the agents embodiment (see Tables III and IV) that reflect the state of the agent's body. They range from a minimum to a maximum values, centered on an ideal value. Physiological state is affected by the agent's interaction with the environment (metabolism and objects).

E. Drives and Motivations

The Drives define the current and past body states that drive the agent attention towards specific needs. They are controlled only by the agent body, which reacts to its environment, with no other interference. They translate physiological changes into specific alarms or urges to action (e.g hunger if blood

Physiological Data	Range	Variation
Blood Sugar	10 - 30 - 50	metabolism: $K_{BSugDec}$*speed food: $K_{BSugInc}$*FoodValue
Endorphine	0 - 20 - 40	metabolism: K_{EndInc} toy: K_{EndDec}*ToyValue
Energy	100 - 120 - 140	metabolism: K_{EnDec}*speed bed: K_{EnInc}*RestValue
Vascular Volume	5 - 25 - 45	metabolism $K_{VVolDec}$*speed water: $K_{VVolInc}$*WaterValue
Pain	0 - 20	metabolism $K_{PainDec}$ obstacles: $K_{PainInc}$*speed

TABLE III
PHYSIOLOGICAL DATA, DRIVES, AND THEIR DYNAMICS.

Constant	Value
$K_{BSugDec}$	-0.0050
$K_{BSugInc}$	0.50
K_{EndDec})	-0.50
K_{EndInc}	0.0020
K_{EnDec}	-0.0030
K_{EnInc})	0.50
$K_{VVolDec}$	-0.0080
$K_{VVolInc}$	0.50
$K_{PainDec}$	-0.005
$K_{PainInc}$	0.50

TABLE IV
PHYSIOLOGICAL VARIABLES CONSTANT VALUES

sugar is low). They vary from a minimum of -10, and to a maximum of 10. The value indicates the excess or absence of certain stimuli in the body, specifically certain physiological needs (by excess or deficit). In any moment the agent can be hungry or not hungry, tired or energetic, etc. This takes in consideration the existence of a "temporary memory": in each moment the drive contains physiological information from the last M (see Equations 4 and 5) time steps. M corresponds to the agent' memory size, where past states influence is attenuated, while ΔV_i corresponds to the deviation of the drive value from its homeostatic position (value between -10 and 10).

$$DriveValue(t) = K * \sum_{i=0}^{M} \left(\Delta V_i * K_i \right) \quad (4)$$

$$K_i = \frac{1}{i+2}; K = \frac{1}{M+2} \quad (5)$$

For instance, a growing level of Blood Sugar level during a certain period, will increase the Hunger drive. This implies a change on the body state (see Fig. 3), and consequent effect on other neural process (e.g. decision making), since they share the same neural network.

As defined in Sec. II, changes in the body state are caused by interaction with objects. But, as shown in Table III, the environment also changes the body of the agent indirectly because of its metabolism. Agent's ongoing tasks (when not interacting with objects) change its internal physiological data; i.e., corresponding to a decrease/increase in a physiological variable according to the metabolism (decrease blood sugar, increase pain, etc.). At each iteration, each Drive is feed and propagated into the NN together with retina signals.

To express its desire to act in the environment, the agent possesses a set of Motivations. These correspond to the level of will to adopt a certain behaviour (Eat, Drink, etc.). The Motivational System is controlled by the neural process. One action is chosen from the Motivations set (neural network output layer), according to a *roulette* algorithm: Motivations with higher value (higher output neurons activation), have more probability of being chosen (see Fig. 3).

F. Background Emotions

The agent's emotional state is processed in parallel and is mirrored into a set Background Emotions (see Sec. I-C). Background Emotions (see Table V) are obtained from the analysis of the Goal System (See Table VI), and the body. The goal system corresponds only to self-survival tasks, related with body state evolution. They reflect the success or failure of a certain self-survival task or behaviour.

Background Emotion	Affected by
wellness/malaise	survivalGoal
relaxation/tension	pain, survivalGoal
fatigue/excitement	energy, ongoingGoal

TABLE V
BACKGROUND EMOTIONS.

Goals	Description
survivalGoal	survival status (0%-100%, low fitness-high fitness)
ongoingGoal	successfully achieved tasks

TABLE VI
GOALS.

G. Reward and Punishment: "feeling" the interaction

Learning is performed by means of an algorithm inspired by the TD-Learning technique, a type of reinforcement learning algorithm [31]. One characteristic of this algorithm is that it associates a Q-value (the predicted reward) with each output continuously, corresponding to the desirability of choosing that behaviour. The reward received in the future (when the task finishes, successfully or not) is used to update the weights that activated the chosen outputs by means of a back-propagation algorithm.

As introduced in Sec. I-C, Emotions are usually associated with either pleasant or unpleasant feelings that can act as a reinforcement [32][28]. For that we created a different reward scheme. The outputs (see Fig. 3) do not have an associated Q-value. Rewards depend on the agent's actions: they are proportional to their effect on the agent's well-being, and their valence (positive or negative) depends on the pleasantness of the new body state. These rewards are received after

interacting with the world, reflecting the consequences of the action performed.

There are two types of rewards: one for the movements and another for the chosen motivation. They are used to update the weights of the NN using a back-propagation algorithm. The agent's internal representation of the objects is an important aspect of the model. The learning process enables the agent to attribute meanings to the objects. They don't have any internal explicit representation.

The following steps are taken in a complete cycle of the system : the new body states are calculated according to the interactions (objects and metabolism), new body state is compared with the previous body state (e.g. pain increase). Goals and Background Emotions are updated, reward is calculated according to the arousal and valence of the interaction (huge pain augment implies negative reward - high arousal, negative valence), the neural network inputs (retina and drives) are refreshed, and finally the neural network outputs are calculated (and new Motivation and movement chosen).

IV. Simulation Results

With this framework we aim, at this point, to test three main hypothesis:

1) An emotional system can emerge from the interaction of self-regulatory Homeostatic Processes and the Environment;
2) The regulation of the Homeostatic Processes influences world categorization and decision making by attributing emotional meanings to objects, and by affecting cognitive processes (e.g. driving attention);
3) Emotions act as a system of internal rewards, that preserve the system, and permit continuous adaptation process in self-survival tasks, by signalling and scaling pleasant or unpleasant interactions/stimuli.

We analyzed several simulations. Each simulation ran with different initial weights of the NN, which were randomly generated. One agent was inserted in a world populated with the objects as referred in Sec.II. Each simulation ran for 40000 cycles. (Each cycle corresponds to an agent's time step to update internal variables, apply Reinforcement Learning, receive and process stimuli, propagate data through the NN, and choose the next action). The following data corresponds to a representative set from our experiments.

A. Adaptation and Self-survival

This section analyzes the agent's ability to regulate its Homeostasis. We assessed the degree of adaptation (Fitness) of the agent, through a Fitness[3] function that expresses the body state of the agent: the agent's *health coefficient*.

$$fitness = 1 - \left(\frac{1}{n * DriveMaxLevel} \right) \cdot \sum_{k=1}^{n} \cdot |drives_i| \quad (6)$$

[3]We use the Fitness concept due to the fact that we are already using Genetic Algorithms in our ongoing work.

DriveMaxLevel stands for the absolute maximum level that Drives can have (10, in the current version), while n corresponds to the current number of Drives (five) and $drives_i$ to the value of each drive (numbered from one to five).

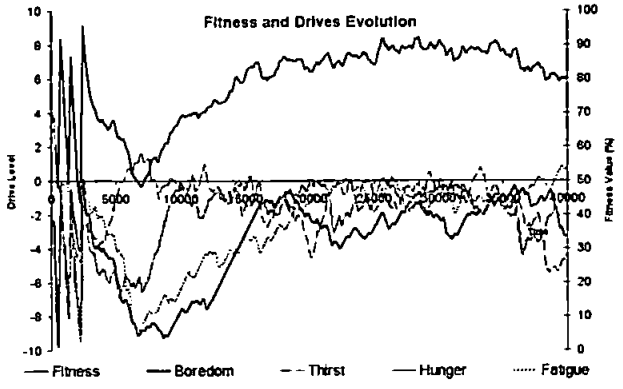


Fig. 4. Fitness and Drives evolution in time: representative simulation 1

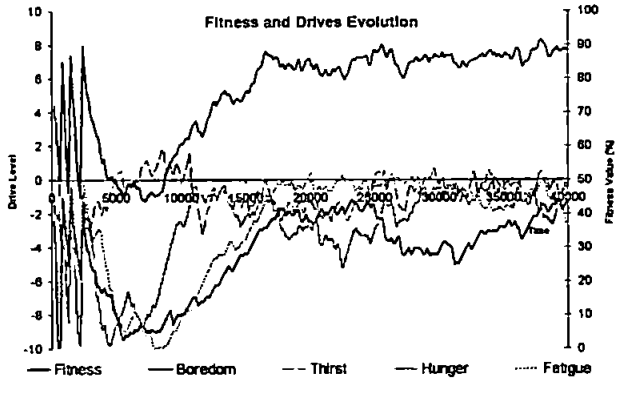


Fig. 5. Fitness and Drives evolution in time: representative simulation 2

Figs. 4 and 5 show the relation between the Fitness function and the evolution of Drives for two representative simulations.

The evolution of the Fitness value in time shows an overall increase of the agent's ability to regulate its body state by interacting with the world. In fact, it can be seen in both charts that after an initial unstable phase the agents were capable of associating the resources in the world with their own internal needs, and use these resources as they needed. Another interesting phenomena can be seen by analyzing the Drives variation in time: when learning allows the agent to reach a stable situation, the Drives variation is maintained within a range of values near to the optimal value (i.e., 0). The agent is not only capable of increasing its Fitness, but it does so maintaining a "healthy behaviour". Extreme body states are avoided, showing the ability of the agent to regulate its own body status, by coping with its metabolism and managing competitive internal stimuli. Note that, as an initial help for the system to learn, when the fitness value is below 40%, the

agent's physiological data are reset to their initial values. This explains the initial picks in the first iterations on the chart.

These results are coherent with our hypothesis. In fact, the properties of an emotional system, can emerge from the interaction between an organism self-regulatory Homeostatic Processes and its Environment (the fundamental role of the body), through the learning process used. The attribution of emotional meanings to objects, through Associative Learning driven by the body, proved to be effective and fundamental to adaptation process.

B. Categorization

We analysed the system further in order to better understand the dynamics of the NN and the learning process. Specially, we wanted to study the effect of the reward system. We suggested previously that an agent would be able to categorize objects in the external world by giving meanings to objects in relation to their body state: the emotional categorization of objects. For these tests we analysed the hidden layer activations of the agent presented in Fig. 4, using Principal Components Analysis (PCA).

To test the NN categorization process, we presented the agent one object at a time. For each object we activated all Drives to their maximum level (also one at the time, see Table VII). For instance, in a state of Hunger, all five objects in the environment were perceived individually. The clusters seen on the PCA chart (Fig. 6) group the stimuli for each object presented to the agent in the tests phase.

Input Number	Object
1-5	Food (Red)
6-10	Bed (Green)
11-15	Obstacle (Blue)
16-20	Water (White)
21-25	Toy (Purple)

TABLE VII
NN OBJECTS STIMULI.

It can be seen that the agent was able categorize the world. In fact, identical external stimuli (objects) are represented internally in a specific and dedicated way.

But how does the agent contextualize the object with its body state? When does it decide to interact with objects in order to survive? Taking a deeper look at these clusters, we can find patterns that are identical to each other. Our hypothesis is that they may represent a second categorization level. A new test scenario was created to analyze these important aspects: in the presence of related object, we varied each of the agent's drives from its minimum value (-10) to its maximum value (+10), considering steps of 2 units (-10, -8, ... 8, 10). See Table VIII. In Fig. 7 we plotted the PCA analysis for the 1st and 2nd Principal Components of this test case, using the object Food, and the drive Hunger.

There is a clear definition of the different drive levels: starting from the extreme need of food (measures 1-4), passing through the absence of the hunger stimulus (measure 6), to the

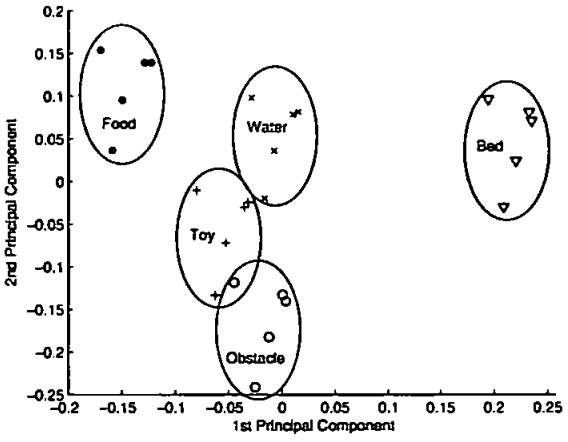


Fig. 6. PCA: Categorization Process

Input Number	Active Drive
1, 6, 11, 16, 21	Hunger
2, 7, 12, 17, 22	Thirst
3, 8, 13, 18, 23	Boredom
4, 9, 14, 19, 24	Fatigue
5, 10, 15, 20, 25	Withdraw

TABLE VIII
NN DRIVES STIMULI.

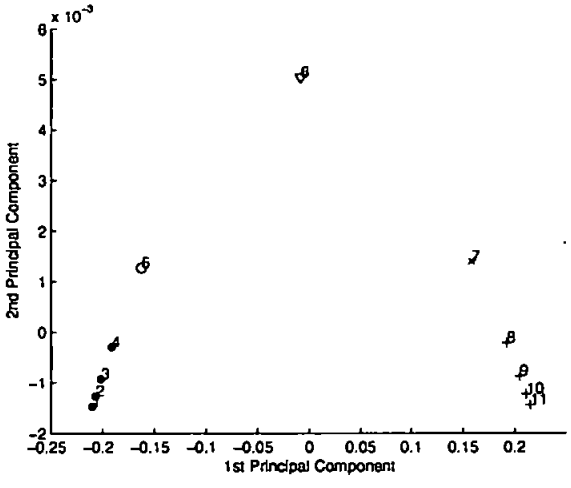


Fig. 7. PCA: variable Hunger level internal representations.

representation of excess of blood sugar (measure 8-11). The agent can identify its own body needs and attribute dynamical meanings to the objects by a 2nd level categorization. It is evident by the sequence the complete separation of the different states of well-being (over-stimulated, homeostatic level, under-stimulated). These results are aligned with our Hypothesis (3): the signalling and scaling of pleasant or

unpleasant interactions/stimuli .

C. The Role of the Body

Next, we tested the detection of "emotional-competent-stimulus" [1], as defined by Damasio. These would be objects or situations (present or remembered) that would lead to a specific emotional state, which could be observable in a stable emotional system.

We considered 3 test cases:

1) No body stimuli (only visual stimuli);
2) No visual stimuli (only body stimuli);
3) Both.

In Fig. 8 drives were kept at zero level (Homeostatic Regime) and we presented each type of object to the agent (no body stimuli) - test case 1. Then, each drive was activated at its maximum level, and the agent was isolated from the environment (no visual stimuli) - test case 2.

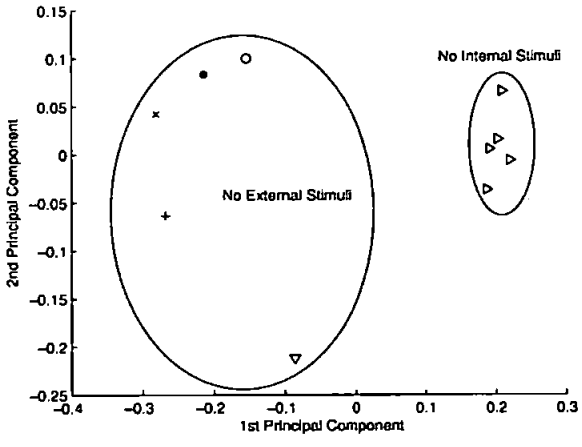


Fig. 8. PCA: test cases.

In Fig. 9 we plot the previous test plus additional more test cases (11 to 15 in the chart): we activated each drive to its maximum level and presented the object that would satiate that need of the agent (e.g. high hunger level in the presence of food) - test case 3.

From both figures, it can be seen that when the agent was situated in its Homeostatic Regime (all drives at level 0, test case 1), objects were categorized within an emotional meaning, not showing special distinction between them. The variance values for each test case can be seen in Table IX [4]. By comparison with the other test cases, test case 1 variance, can be considered very low. This seems coherent with our expectancies, taking into account that the objects have a meaning when related with the body. Even though a discrimination among objects can be seen: probably an influence from the new neural predisposition (after interacting and learning with the experience).

[4]The variance was calculated for the different test cases using the first 3 Principal Components. The presented values for the variance are separated by dimension.

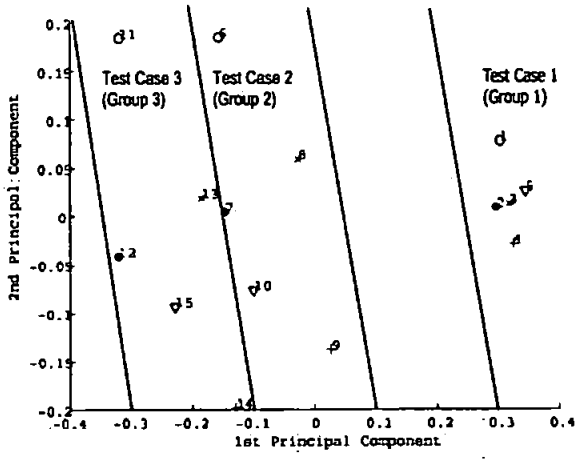


Fig. 9. PCA: test cases.

Test Case	Component 1 (x)	Component 2 (y)	Component 3 (z)
1	0.0004	0.0015	0.0026
2	0.0062	0.0156	0.0144
3	0.0071	0.0201	0.0127

TABLE IX
VARIANCE VALUES FOR THE FIRST 3 PRINCIPAL COMPONENTS (DIMENSIONS).

In the absence of external stimuli (test case 2) the agent identified clearly its body state, and its body needs trough the creation of an implicit body map (i.e., a internal representation of the body state). This fact becomes even more evident in test case 3, when the object to satiate the body need is present (Fig. 9): objects and drives are strongly associated, showing that the agent acquired specific ways to respond to specific internal and external events (similar symbols refer to the same Drive).

Distance	Value
d_{31}	0.4092
d_{32}	0.0365
d_{21}	0.3754

TABLE X
DISTANCES BETWEEN CLUSTERS' ATOMIC POINTS.

Summarizing, several facts can be observed: the variance in test case 1 ("*no-internal-stimuli*") is lower than in both other test cases; test case 3 results are closer to test case 2 results ("*no-external-stimuli*"), than to test case 1 (see Table X): $d_{31} = 0.4092$ (distance from groups 3 to 1), and $d_{32} = 0.0365$ (distance from group 3 to 2); group 1 presents a similar pattern to the other groups, an influence in the categorization process.

These observations indicate a great influence of the body on the cognitive processes, in the case related with the internal representation of a perceived object. Indicates also the preferential perceptual processing regarding the pleasant or

unpleasant "meaning" of objects, influencing the accuracy of object representation during the perceptual process. There is a scaling factor for the object's internal representation (as seen in the previous section). These results are strongly coherent with Damasio's theory [1].

V. Conclusion and Future Work

Damasio refers to the importance of Emotions in assisting an individual to maintain its survival because they seem to be an important mechanism for adaptation and decision making in dynamical systems [1], [16], [17]. In this phase of our work we focused on the basic foundations of Emotions from an evolutionary perspective: we assumed the existence of neural pathways that facilitate survival. Moreover we use the dimensional approach (arousal/valence), together with an Associative Learning process [28], to drive adaptation contingencies, using the body to drive such process.

We addressed the notion of the emergence of a stable emotional system by means of self-regulatory Homeostatic Processes. In the previous section we demonstrated that it is possible to model such phenomenon. As suggested by Damasio [1], environmental events of value should be susceptible to preferential perceptual processing regarding their pleasant or unpleasant meaning. We believe that the architecture and specially the reward system (the agent's appetite for well-being) were responsible for the emergence of stable emotional systems in our simulations. Furthermore, the results are coherent with Damasio's convincing theories about the existence of a background emotional system [1]. We demonstrated that phenomena such as body/world categorization and existence of a body map can evolve from a simple rule: self-survival. As already discussed in the previous sections, we were able to evaluate our hypothesis.

Our model also demonstrated that physical restrictions (even with a very simple artificial embodiment) can play an important role in the adaptation of the agent to its environment. The use of a learning algorithm based on the environment and embodiment allows for the agent's "brain" to dynamically categorize the world regarding bodily, environmental and individual aspects (metabolism). Agent and environment are strongly coupled in learning and living. The emergence of a stable emotional system (albeit in low level tasks), potentiated dynamical categorization of objects due to their emotional context, proving to be effective and versatile enough to allow the agent to adapt to an unknown environment.

Currently, we are looking at more complex tasks to be performed on top of our background emotional model. We are in the process of defining a system of foreground emotional states, and developing the current one. A more close investigation about the changes in the body state due to an induced emotion, is also an interesting perspective. At this stage we will then be able to develop our investigations on the role of music in emotional states, and on the possible existence of co-evolutionary mechanisms reinforcing the relation between Emotions and Music.

On the long run, we hope to apply our model to decision making tasks (e.g. music composition), as it allows to reduce the space state of choices, through an emotional categorization. Another interesting perspective comes from recent claims, specially in Robotics, more specifically in a new field: Internal Robotics [33].

Acknowledgment

The authors would like to acknowledge the financial support of the Portuguese Foundation for Science and Technology (FCT, Portugal).

References

[1] A. Damasio, *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. Vintage, 2000.
[2] C. Darwin, *The Expression of the Emotions in Man and Animals*, P. Ekman, Ed. Oxford University Press, 1998.
[3] W. James, "What is an emotion?" *Mind*, vol. 9, pp. 188–205, 1884. [Online]. Available: http://psychclassics.yorku.ca/James/emotion.htm
[4] W. Wundt, *Outlines of Psychology*. Wilhelm Engelmann, 1897.
[5] R. Lazarus, *Emotion and Adaptation*. USA: Oxford University Press, 1991.
[6] W. Cannon, *Bodily Changes in Pain, Hunger, Fear and Rage*. New York: Appleton, 1929.
[7] S. S. Tomkins, "Affect, imagery, consciousness," *The positive affects*, vol. 1, 1962.
[8] ——, "The role of facial response in the experience of emotion," *Journal of Personality and Social Psychology*, vol. 40, pp. 351–357, 1981.
[9] R. Plutchik, "Emotion: Theory, research, and experience," *Theories of emotion*, vol. 1, pp. 3–33, 1980.
[10] ——, *The Emotions*. University Press of America, 1991.
[11] C. E. Izard, *The face of emotion*. New York: Appleton-Century-Crofts, 1971, vol. 1.
[12] ——, *Human Emotions*. Plenum Press, 1977.
[13] ——, "Four systems for emotion activation: Cognitive and noncognitive processes," *Psychological Review*, vol. 100, pp. 68–90, 1993.
[14] P. Ekman, "Basic emotions," in *The Handbook of Cognition and Emotion*, T. Dalgleish and T. Power, Eds. Sussex, U.K.: John Wiley and Sons, Ltd., 1999, pp. 45–60.
[15] ——, *Darwin and facial expression: A century of research in review.*, P. Ekman, Ed. Academic, 1973.
[16] A. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. Avon books, 1994.
[17] ——, *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*. Harcourt, 2003.
[18] E. T. Rolls, *The Brain and Emotion*. Oxford University Press, 1999.
[19] J. Panksepp, "The neuro-evolutionary cusp between emotions and cognitions: Implications for understanding consciousness and the emergence of a unified mind science," *Consciousness & Emotion*, vol. 1, no. 1, pp. 15–54, 2000.
[20] R. J. Dolan, "Emotion, cognition, and behavior," *Science Magazine*, vol. 298, pp. 1091–1094, November 2002.
[21] J. A. Russel, *Emotion: Theory, research, and experience*. Toronto: Academic, 1989, vol. 4, ch. Measures of emotion, pp. 83–111.
[22] L. A. Feldman, "Valence-focus and arousal-focus: Individual differences in the structure of affective experience," *Journal of Personality and Social Psychology*, vol. 69, pp. 153–166, 1995.
[23] R. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 23, p. 11751191, 2001.
[24] J. D. Velasquez, "Modeling emotion-based decision-making," in *Proceeding of 1998 AAAI Fall Symposium Emotional and Intelligent: The Tangled Knot of Cognition (Technical Report FS-98-03)*. Orlando, FL: AAAI Press, 1998, pp. 164–169.
[25] D. Canamero, "A hormonal model of emotions for behavior control," in *4th European Conference on Artificial Life ECAL'97*, 1997.
[26] C. Breazeal, "Emotions and sociable humanoid robots," *International Journal Human-Computer Studies*, vol. 59, pp. 119–155, 2003.
[27] S. C. Gadanho and J. Hallam, "Robot learning driven by emotions," *Adaptive Behavior*, vol. 9, pp. 42–64, 2001.

[28] E. T. Rolls, "Memory systems in the brain." *Annu. Rev. Psychol.*, vol. 51, pp. 599–630, 2000.
[29] E. Coutinho, H. Pereira, A. Carvalho, and A. Rosa, "Livia - life, interaction, virtuality, intelligence, art," 2003, an Ecological Simulator Of Life.
[30] N. Gracias, H. Pereira, J. A. Lima, and A. Rosa., "An artificial life environment for ecological systems simulation." in *Proc. A-Life V Conference, ALIFE V'96*, 1996.
[31] R. S. Suton and A. G. Barto, *Reinforcement Learning: an introduction*, ser. Adaptive Computation and Machine Learning, T. Dietterich, Ed. Cambridge (MA), USA: MIT Press, 2002.
[32] S. S. Tomkins, "Affect theory," in *Approaches to emotion*, K. R. Scherer and P. Ekman, Eds. Hillsdale (NJ), USA: Erlbaum, 1984.
[33] D. Parisi, "Internal robotics," *Connection Science*, vol. 16, no. 4, pp. 325–338, December 2004.

Coutinho, E., Gimenes, M., Martins, J., & Miranda, E. (2005). Computational musicology: An artificial life approach. *In Bento, C., Cardoso, A. & Dias, G. (Eds.), Proceedings of the Portuguese Conference on Artificial Intelligence.* Covilhã (Portugal): IEEE Press, pp. 85-93 .

Computational Musicology: An Artificial Life Approach

Eduardo Coutinho, Marcelo Gimenes, João M. Martins and Eduardo R. Miranda
Future Music Lab
School of Computing, Communications & Electronics
University of Plymouth
Drake Circus
Plymouth PL4 8AA
UK
Email: {eduardo.coutinho,marcelo.gimenes,joao.martins,eduardo.miranda}@plymouth.ac.uk

Abstract— **Artificial Life (A-Life) and Evolutionary Algorithms (EA) provide a variety of new techniques for making and studying music. EA have been used in different musical applications, ranging from new systems for composition and performance, to models for studying musical evolution in artificial societies. This paper starts with a brief introduction to three main fields of application of EA in Music, namely *sound design*, *creativity* and *computational musicology*. Then it presents our work in the field of computational musicology. Computational musicology is broadly defined as the study of Music with computational modelling and simulation. We are interested in developing A-Life-based models to study the evolution of musical cognition in an artificial society of agents. In this paper we present the main components of a model that we are developing to study the evolution of musical ontogenies, focusing on the evolution of rhythms and emotional systems. The paper concludes by suggesting that A-Life and EA provide a powerful paradigm for computational musicology.**

Index Terms — **Artificial Life, Music, Computational Musicology, Synchronization and Evolution of Rhythm, Musical Ontogensis, Emergence of Emotion.**

I. Introduction

Acoustics, Psychoacoustics and Artificial Intelligence (AI) have greatly enhanced our understanding of Music. We believe that A-Life and EA have the potential to reveal new understandings of Music that are just waiting to be unveiled.

EA have varied applications in Music, with great potential for the study of the artificial evolution of music in the context of the cultural conventions that may emerge under a number of constraints, including psychological, physiological and ecological constraints.

We identify three main fields of application of EA in Music: *sound design*, *creativity* and *computational musicology*. The following sections briefly survey these three main fields of application. Then we introduce our work in the field of computational musicology, inspired on A-Life techniques and EA.

A. Sound Design

The production of sound faced a revolution in the middle of the 20^{th} century with the appearance of the digital computer [1]. Computers were given instructions to synthesise new sounds algorithmically. Synthesisers (or *software synthesisers*) soon became organized as a network of functional elements (signal generators and processors) implemented in software. Comprehensive descriptions of techniques for computer sound synthesis and programming can be found in the literature [2].

The vast space of parameter values that one needs to manage in order to synthesise sounds with computers led many engineers to cooperate with musicians in order find effective ways to navigate in this space. Genetic algorithms (GA) have been successfully used for this purpose [3]. EA have also been used to develop topological organizations of the functional elements of a software synthesiser, using Genetic Programming (GP) [4].

The use of extremely brief time-scales gave rise to granular synthesis [5], a technique that suits the creation of complex sounds [6], adding more control problems to the existing techniques. One of the earliest applications of EA to granular synthesis is *Chaosynth*, a software designed by Miranda [7] that uses Cellular Automata (CA) to control the production of sound grains. *Chaosynth* demonstrates the potential of CA for the evolution of oscillatory patterns in a two-dimensional space. In most CA implementations, CA variables (or cells) placed on a 2D matrix are often associated with colours, creating visual patterns as the algorithm evolves in time. However, in *Chaosynth* the CA cells are associated with frequency and amplitude values for oscillators. The amplitude and frequency values are averaged within a region of the 2D CA matrix, corresponding to an oscillator. Each oscillator contributes a partial to the overall spectrum of a grain. The spectra of the grains are generated according to the evolution of the CA in time (Fig. 1).

More recently, Mandelis and Husbands [8] developed *Genophone*, a system that uses genetic operators to create new generations of sounds from two sets of preset synthesis parameters. Some parameters are left free to be manipulated with a data-glove by an external user, who also evaluates the fitness of the resulting sounds. Offspring sounds that are ranked best by the user will become parents of a new population of sounds. This process is repeated until satisfactory sounds are found.

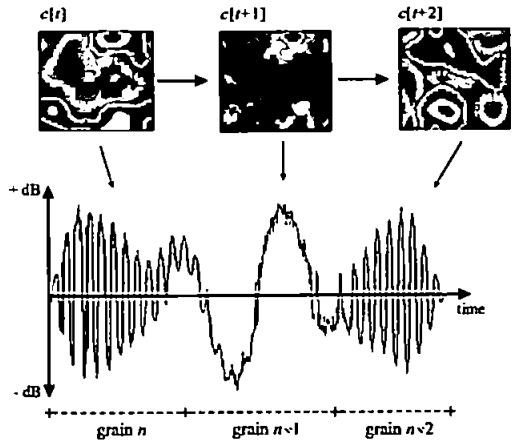


Fig. 1. Each snapshot of the CA produces correspond to a sound-grain. (Note, however, that this is only a schematic representation, as the grains displayed here do not actually correspond to these particular snapshots.)

B. Creativity

One interesting question with respect to the use of computers for aiding musical creativity is whether computers can create new kinds of musical compositions. In this case, the computer should neither be embedded with particular well-known compositional models at the outset nor learn from selected examples, which is not the case with most Artificial Intelligence-based systems for generating musical compositions.

Composers have used a number of mathematical models such as combinatorial systems, grammars, probabilities and fractals [9][10][11] to compose music that does not imitate well-known styles. Some of these composers created very interesting pieces of new music with these models and opened innovative grounds in compositional practices, e.g., the techniques created by Xenakis [12].

The use of the emergent behaviour of EA, on the other hand, is a new trend that is becoming very popular for its potential to generate new music of relatively good quality. A great number of experimental systems have been used to compose new music using EA: Cellular Automata Music [13], CA Music Workstation [14], CAMUS [15], MOE [16], GenDash [17], CAMUS 3D [18], Living Melodies [19] and Genophone [20], to cite but a few.

For example, *CAMUS* [15] takes the emergent behaviour of Cellular Automata (CA) to generate musical compositions. This system, however, goes beyond the standard use of CA in music in the sense that it uses a two-dimensional Cartesian representation of musical forms. In this representation the coordinates of a cell in the CA space correspond to the distances between the notes of a set of three musical notes.

As for GA-based generative music systems, they generally follow the standard GA procedures for evolving musical materials such as melodies, rhythms, chords, and so on. One example of such system is Vox Populi [21], which evolves populations of chords of four notes, through the operations of crossover and mutation.

EA have also been used in systems that allow for interaction in real-time; i.e., while the composition is being generated. In fact, most GA-based systems allow for this feature by letting the user to control GA operators and fitness values while the system is running. For example, Impett proposed an interesting swarm-like approach to interactive generative musical composition [22]. Musical composition is modelled here as an agent system consisting of interacting embodied behaviours. These behaviours can be physical or virtual and they can be emergent or preset. All behaviours co-exist and interact in the same world, and are adaptive to the changing environment to which they belong. Such behaviours are autonomous, and prone to aggregation and generation of dynamic hierarchic structures.

C. Computational Musicology

Computational musicology is broadly defined as the study of Music by means of computer modelling and simulation. A-Life models and EA are particularly suitable to study the origins and evolution of music. This is an innovative approach to a puzzling old problem: if in Biology the fossils can be studied to understand the past and evolution of species, these "fossils" do not exist in Music; musical notation is a relatively recent phenomenon and is most prominent only in the Western world. We are aware that Musicology does not necessarily need computer modelling and simulation to make sense. Nevertheless, we do think that "*in silico*" simulation can be useful to develop and demonstrate specific musical theories. These theories have the advantage that they can be objective and scientifically sound.

Todd and Werner [23] proposed a system for studying the evolution of musical tunes in a community of virtual composers and critics. Inspired by the notion that some species of birds use tunes to attract a partner for mating, the model employs mating selective pressure to foster the evolution of fit composers of courting tunes. The model can co-evolve male composers who play tunes (i.e., sequences of notes) along with female critics who judge those songs and decide with whom to mate in order to produce the next generation of composers and critics. This model is remarkable in the sense that it demonstrates how a Darwinian model with a pressure for survival mechanism can sustain the evolution of coherent repertoires of melodies in a community of software agents. Miranda [24] [25] proposed a mimetic model to demonstrate that a small community of interactive distributed agents furnished with appropriate motor, auditory and cognitive skills can evolve from scratch a shared repertoire of melodies (or tunes) after a period of spontaneous creation, adjustment and memory reinforcement. One interesting aspect of this model is the fact that it allows us to track the development of the repertoire of each agent of the community. Metaphorically, one could say that such models enable us to trace the musical development (or "education") of an agent as it gets older.

From this perspective we identify three important components of an Artificial Musical Society: agents synchronization, knowledge evolution, and emotional content in performance.

The first presents itself as the basis for musical communication between agents. The second, rooted on the first, allows musical information exchange, towards the creation of a cultural environment. Finally we incorporate the indispensable influence of emotions in the performance of the acquired music knowledge. The following sections present this three aspects separately. Even though they are parts of the same model, experiments were run separately. We are working towards the complete integration of the model, and co-evolution of the musical forms: from motor response to compositional processes and performances.

II. Emergent Beat Synchronisation

A. Inspiration: Natural Timing

Agents interacting with one another by means of rhythm need mechanisms to achieve beat synchronisation.

In his book *Listening*, Handel [26] argues that humans have a biological constrain referred to as *Natural Timing* or *Spontaneous Tempo*. This means that when a person is asked to tap an arbitrary tempo, they will have a preference. Furthermore, if the person is asked to tap along an external beat that is faster or slower, and if the beat suddenly stops, then they will tend to approximate to their preferred tempo. The tap interval normally falls between 200 msec and 1.4 sec, but most of the tested subjects were in the range of 200 - 900 msec [27]. The claim that this phenomenon is biologically coded rises from the extreme proximity of these values when observed in identical twins. The same disparity observed for unrelated subjects is observed in fraternal twins. The time interval between two events is called Inter-Onset Interval (IOI).

In our model, the agents "are born" with different natural timings by default. As they interact with each other, each agent adapts its beat to the beats of the other agents.

B. Synchronisation Algorithm

Computational modeling of beat synchronisation has been tackled in different ways. Large and Kolen devised a program that could tap according to a rhythmic stimulus with nonlinear-oscillators [28], using the gradient descendant method to update their frequency and phase. Another approach, by Scheirer, consisted of modelling the perception of meter using banks of filters [29]. We propose an algorithm based on Adaptive Delta Pulse Code Modulation (ADPCM) that enables the adaptation of different agents to a common ground pulse, instead of tracking a given steady pulse. Our algorithm proved to be more compatible with Handel's notion of natural timing, as discussed in the previous section. As in ADPCM for audio, where a variable time step tracks the level of an audio signal, the agent in our model uses a variable time step to adjust its IOI to an external beat. The agent counts how many beats from the other agents fit into its cycle and it determines its state based on one of the following conditions: SLOW (listened to more than one beat), FAST (no beats were listened), or POSSIBLY_SYNCHRONISED (listened to one beat). Depending on whether the agent finds itself in one of the first two states, it increases or decreases the size of their IOIs. Delta corresponds to the amount by which the value of an IOI is changed. If the agent is in the POSSIBLY_SYNCHRONISED state and the IOIs do not match, then there will be a change of state after some cycles, and further adjustments will be made until the IOIs match. However, the problem is not solved simply by matching the IOI of the other agent. Fig. 2(b) illustrates a case where the IOIs of two agents are the same but they are out of phase. An agent solves this problem by delaying its beat until it produces a beat that is close to the beat of the other agent (Fig. 2(c)).

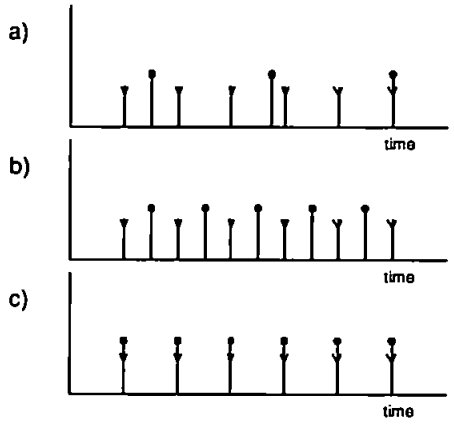


Fig. 2. (a) The agents have different IOIs; (b) The agents have the same IOI but they are out of phase; (c) The IOIs are synchronised.

C. Experiment and Result

In this section we present the result of an experiment with two agents adapting to each other's beats. Fig. 3 shows the temporal evolution of the IOIs of the agents. The minimum value for Delta, which is also the initial value of the time step, is different for the two agents. If the agent recognises that it is taking too long to change its state, the former value of Delta is multiplied by 2. Oscillatory patterns were observed when they were close to finding a common beat, due to the fact that both agents changed their IOIs and phases when they realised that they were not synchronised. The solution to this problem was solved by letting only one of the agents to change the phase after hearing one beat from the other agent.

Agent 1 started with an IOI equal to 270 ms and it had an initial adaptation step of 1 ms. Agent 2 started with an initial IOI equal to 750 ms and it had an initial adaptation step of 3 ms. Fig. 3 shows that the agents were able to find a matching IOI of 433 ms and synchronise after 26 beats. Notice that they found a common IOI after 21 beats, but they needed 5 more beats to synchronise their phases.

One interesting alternative that requires further study is the interaction between the agents and a human player. In the present case study the system requires many beats to reach synchronisation, but it is expected that the ability that humans have to find a common beat quickly may introduce a shortcut into the whole process.

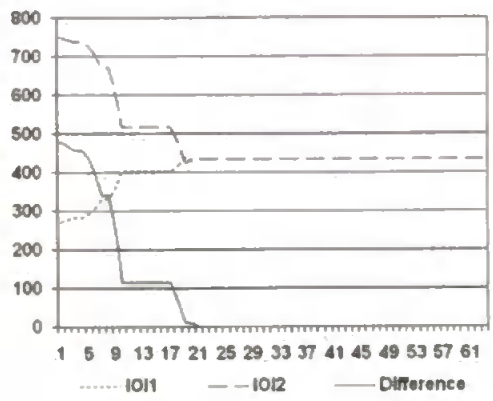


Fig. 3. Evolution of IOIs and their difference.

In this experiment, the "spontaneous tempo" and the Delta values of the agents were initialised by hand. But once the synchronisation algorithm is embedded in a model to study the evolution of musical rhythm one needs to implement a realistic way to initialise these values. Different agents can be implemented with different default Delta value but it would be more realistic to devise a method to modulate such value in function of some form of musical expression, or semantics. In order to do this, we are looking into ways in which we could program the agents to express emotions. In this case, the agents should be given the ability to modulate Delta coefficients and initial deviations from their "spontaneous tempo" in function of their emotional state. In section IV we present the first phase of an emotional system that we are developing to implement this.

III. Musical Ontogenesis in an Artificial Society

In Philosophy of Science, ontogenesis refers to the sequence of events involved in the development of an individual organism from its birth to its death. We therefore use the term musical ontogenesis to refer to the sequence of events involved in the development of the musicality of an individual. Intuitively, it should be possible to predict the music style of future musicians according to restrained music material that is absorbed during their formative stages. But would it be possible to objectively study the way in which composers or improvisers create music according to their educational background? Although it may be too difficult to approach this subject with real human musicians, we suggest that it should be possible to develop such studies with artificial musicians. A model of musical ontogenesis is therefore useful to study the influence of the musical material learned during the formative years of artificial musicians, especially in systems for musical composition and improvisation. A growing number of researchers are developing computer models to study cultural evolution, including musical evolution ([30] [31] [32] [33]). Gimenes [34] presents RGeme, an artificial intelligence system for the composition of rhythmic passages inspired by Richard Dawkin's theory of memes. Influenced by the notion that genes are units of genetic information in Biology, memes are defined as basic units of cultural transmission. A rhythmic composition would be understood as a process of interconnecting ("composition maps") sequences of basic elements ("rhythmic memes"). Different "rhythmic memes" have varied roles in the stream. These roles are learned from the analysis of musical examples given to train the system.

A. RGeme

The overall design of the system consists of two broad stages: the learning stage and the production stage. In the learning stage, software agents are trained with examples of musical pieces in order to evolve a "musical worldview". The dynamics of this evolution is studied by analysing the behaviour of the memes logged during the interaction processes.

At the beginning of a simulation a number of Agents is created. They sense the existence of music compositions in the environment and choose the ones with which they will interact, according to some previously given parameters such as the composer's name and the date of composition.

Agents then parse the chosen compositions to extract rhythmic memes (Candidate Memes) and composition maps. The new information is compared with the information that was previously learned and stored in a matrix of musical elements (Style Matrix). All the elements in the Style Matrix possess a weight that represents their relevance over the others at any given moment. This weight is constantly changing according to a transformation algorithm that takes into account variables such as the date the meme was first listened to, the date it was last listened to and a measure of distance that compares the memes stored in the Style Matrix and the Candidate Memes. These features can be seen in more detail in [34].

At last, in the production phase the Agents execute composition tasks mainly through the reassignment of the various Composition Maps according to the information previously stored in the learning phase.

B. Experiment and Result

The different Style Matrices that are evolved in an agent's lifetime represent the evolution of its musical worldview. One can establish the importance of the diversity of the raw material (in terms of developing different musical worldviews) based on the data stored in the Style Matrix's log files. It is possible to directly control the evolution of an agent's worldview, for instance, by experimenting with different sets of compositions originated from different composers.

In Fig. 4 we show the results obtained from an experiment involving a simple learning scenario. During a given period of time an agent only interacted with a particular set of compositions by Brazilian composer Ernesto Nazareth. Afterwards, the agent interacted with a different set of compositions by another Brazilian composer, Jacob do Bandolim. In the same figure, each line represents the evolution of the relative importance (weight) of a small selection of memes that the agent learned during the simulation. Fig. 5 shows the musical notation for

each one of these memes. We can observe different behaviours in the learning curves, which means that the agent was exposed to each one of these memes in different ways.

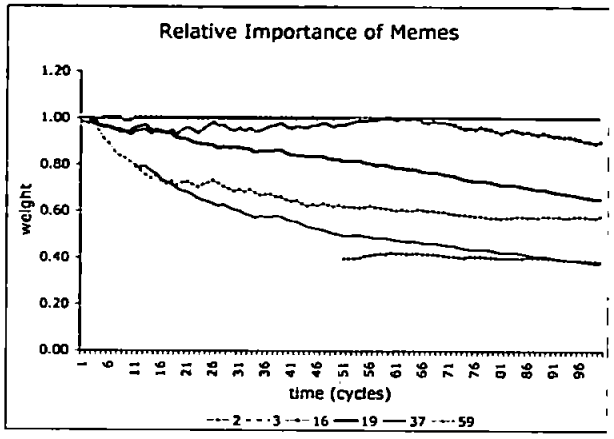


Fig. 4. Relative importance of memes in time.

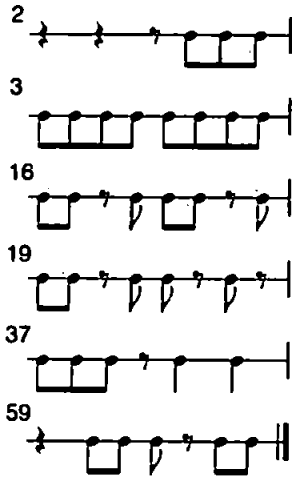


Fig. 5. Musical representation of rhythmic memes.

RGeme has the potential to execute intricate simulations with several Agents learning at the same time from rhythms by composers from inside and outside the system's environment. We believe that this model will allow for the objective establishment of a sophisticated musical ontogenesis through which one will be able to control and predict the musical culture of the inhabitants of artificial communities.

There is however a number of problems that needs to be addressed in order to increase the complexity of this model. One such problem is beat synchronisation, which has been discussed in the previous section. It is possible to observe the behaviour of thousands of male fireflies flashing synchronously during their mating season. Each insect has its own preferred pulse but they gradually adjust their pulses to a single global beat by observing each other [35]. Different humans also have their own preferred pulses, which are driven towards synchrony when engaged in collective musical performance with other humans, non-humans or both. As with fireflies, this mechanism is believed to be biologically coded in humans.

Nonetheless, music is mostly the result of a cultural context [36]. Specially in our research, the rules for composition and performance should emerge from social interactions of agents.

IV. Modelling Emotions

A. Expressivity

The use of expressive marks by Western composers documents well the common assumption that emotions play an important role in music performance.

Expressive marks are performance indications, typically represented as a word or a short sentence written at the beginning of a movement, and placed above the music staff. They describe to the performer the intended musical character, mood, or emotion as an attribute of time, as for example, *andante con molto sentimento*, where *andante* represents the tempo marking, and *con molto sentimento* its emotional attribute.

Before the invention of the metronome by Dietrich Nikolaus Winkel in 1812, composers resorted to words to describe the tempo (the rate of speed) in a composition: Adagio (slowly), Andante (walking pace), Moderato (moderate tempo), Allegretto (not as fast as allegro), Allegro (quickly), Presto (fast). The metronome's invention provided a mechanical discretization of musical time by a user chosen value (beat-unit), represented in music scores as the rate of beats per minute (quarter-note = 120). However, after the metronome's invention, words continued to be used to indicate tempo, but now often associated with expressive marks. In some instances, expressive marks are used in lieu of tempo markings, as previous associations indicate the tempo being implied (e.g. *funebre* implies a slow tempo).

The core "repertoire" of emotional attributes in music remains short. Expressions such as *con sentimento*, *con bravura*, *con affetto*, *agitato*, *appassionato*, *affetuoso*, *grave*, *piangendo*, *lamentoso*, *furioso*, and so forth, permeate different works by different composers since Ludwig van Beethoven (1770-1827) (for an example see Fig. 6) . But what exactly do these expressions mean?

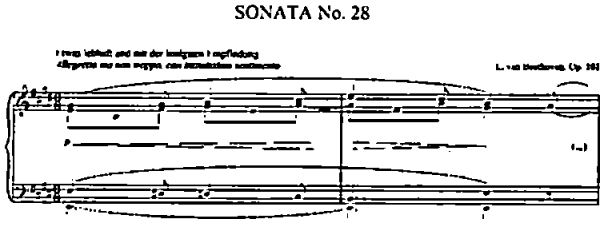


Fig. 6. Beethoven score: example of use of emotional attributes.

Each performer holds a different system of beliefs of what expressions such as *con sentimento* represent, as our understanding of emotions has not yet reduced them to a lawful

behaviour. Without consensus on the individual meaning of such marks, a performance *con sofrimento* is indistinguishable from one *con sentimento*, since both expressions presume an equally slow tempo. Although we have no agreement on the meaning of expressive marks and their direct musical consequences, musicians have intuitively linked expressivity with irregularity within certain boundaries. Celebrated Polish pianist and composer Ignacy Jan Paderewski (1860-1941) stated: "every composer, when using such words as *espressivo*, *con molto sentimento*, *con passione*, and so on, demands (...) a certain amount of emotion, and emotion excludes regularity... to play Chopin's G major Nocturne with rhythmic rigidity and pious respect for the indicated rate of movement would be (...) intolerably monotonous (...). Our human metronome, the heart, under the influence of emotion, ceases to beat regularly - physiology calls it arrhythmic, Chopin played from his heart. His playing was not rational, it was *emotional*" [37].

Composers are well aware that a clear representation of the musical idea reduces ambiguity in the interpretation of the message (the music score). However, the wealth of shadings, accents, and tempo fluctuations found in human performances are, at large, left unaccounted by the composer as the amount of information required to represent these type of nuances carries, in practice, no linear bearing in the detail human performers can faithfully reproduce.

While the electronic and computer music mediums provide composers the power to discretize loudness and time related values in very small increments (for example, MIDI systems [38] use 128 degrees of loudness, and time measured in milliseconds), we note that music scores for human performances use eight approximate levels of loudness (ppp, pp, p, mp, mf, f, ff, fff), and time is discretized in values hundreds of milliseconds long. If we compare any two "faithful" human performances of a work, we conclude that, from performance to performance, only the order of notes remains strictly identical.

Expression marks operate as synesthesia, that is, the stimulation of one sense modality to rise to a sensation in another sense modality [39]. Although their direct musical consequences remain unclear, we can deduce which musical levels are susceptible of being influenced: time and loudness.

These are structural levels where small value changes produce significantly different results. The amount of information needed to describe such detail in fine resolution falls outside the precision limits with which human performers process a music score to control time and the mechanics of traditional music instruments.

> "*Look at these trees!*" Liszt told one of his pupils, "*the wind plays in the leaves, stirs up life among them, the tree remains the same. That is Chopinesque rubato*[1]."

[1] *Rubbato*: from the Italian "robbed", used to denote flexibility of tempo to achieve expressiveness.

B. Emotions

We go back to the 19^{th} century to find the earliest scientific studies: Darwin's observations about bodily expression of emotions [40], James's studies on the meaning of emotion [41], and Wundt's work on the importance of emotions for Psychology [42]. But studies on behaviour focused for many years only on higher level cognitive processes, discarding emotions [43]. Still, emotions were occasionally discussed, and the ideas changed considerably within the last decade or so. Research connecting mind and body, and the role of emotions in rational thinking gained prominence after the work of Cannon and Bard [44]. In short, they suggested that there are parallel neural paths from our senses to the experience of an emotion and to its respective physiological manifestation. Later Tomkins [45][46], Plutchik [47][48] and Izard [49][50][51][52] developed similar theories. They suggested that emotions are a group of processes of specific brain structures and that each of these structures has a unique concrete emotional content, reinforcing their importance. Ekman proposed a set of basic (and universal) emotions [53], based on cross-cultural studies [54]. These ideas were widely accepted in evolutionary, behavioural and cross-cultural studies, by their proven ability to facilitate adaptive responses.

Important insights come from Antonio Damasio [55][56][57], who brought to the discussion some strong neurobiological evidence, mainly exploring the connectivity between body and mind. He suggested that, the process of emotion and feeling are part of the neural machinery for biological regulation, whose core is formed by homeostatic controls, drives and instincts. Survival mechanisms are related this way to emotions and feelings, in the sense that they are regulated by the same mechanisms. Emotions are complicated collections of chemical and neural responses, forming a pattern; all emotions have some regulatory role to play, leading in one way or another to the creation of circumstances advantageous to the organism exhibiting the phenomenon. The biological function of emotions can be divided in two: the production of a specific reaction to the inducing situation (e.g. run away in the presence of danger), and the regulation of the internal state of the organism such that it can be prepared for the specific reaction (e.g. increased blood flow to the arteries in the legs so that muscles receive extra oxygen and glucose, in order to escape faster). Emotions are inseparable from the idea of reward or punishment, of pleasure or pain, of approach or withdrawal, or personal advantage or disadvantage.

Our approach to the interplay between music and emotions follows the work of these researchers, and the relation between physiological variables and different musical characteristics [58]. Our objective is to develop a sophisticated model to study music performance related to an evolved emotional system. The following section introduces the first result of this development.

Physiological Data	Drives	Variation
Adrenaline	Explore	neural activity (arousal)
Blood Sugar	Hunger	metabolism, food
Endorphine	Boredom	metabolism, toys
Energy	Fatigue	metabolism, bed
Vascular Volume	Thirst	metabolism, water
Pain	Withdraw	metabolism, obstacles
Heart Rate	-	metabolism, all objects

TABLE I
PHYSIOLOGICAL DATA, DRIVES, AND THEIR DYNAMICS.

C. The Model

The current version of our model consists of an agent with complex cognitive, emotional and behavioural abilities. The agent lives in an environment where it interacts with several objects related to its behavioural and emotional states. The agent's cognitive system can be described as consisting of three main parts: Perceptual, Behavioural, and Emotional systems.

The Perceptual system (inspired in LIVIA [59] and GAIA [60]) receives information from the environment through a retina modelled as close as possible to a biological retina in functional terms. It senses a bitmap world through a *ray tracing* algorithm, inspired by the notion that photons travel from the light-emitting objects to the retina. The Behavioural system is divided into two sub-systems: Motivational and Motor Control. These sub-systems define the interaction of the agents with their environment. While the agents interact with objects and explore the world, the Motivational sub-system uses a feed-forward neural network to integrate visual input and information about their internal and physiological states. The network learns through a reinforcement learning algorithm. As for the Motor Control sub-system, the agents control their motor system by means of linear and angular speed signals, allowing them to navigate in their world; this navigation includes obstacle avoidance and object interaction. The Emotional system considers the role of emotions as part of an homeostatic mechanism [56]. The internal body state of an agent is defined by a set of physiological variables that vary according to their interaction with the world and a set of internal drives. The physiological variables and the internal drives in the current version of the model are listed in Table I. The agents explore the world and receive the stimuli from it. Motor Control signals are also controlled by the neural network. There are several types of objects: food, water, toys, beds, and obstacles. Each of them is related to one or more physiological variables. Interacting with objects causes changes in their internal body state. For instance, the Vascular Volume (refer to Table I) of an agent will be increased if it encounters water and manifests the desire to drink it. The agent's own metabolism can also change physiological data; e.g. moving around the world decrease the energy level of an agent. An emotional state reflects the agent's well-being, and influences its behaviour through an amplification of its body alarms. For further details on the model, refer to [61].

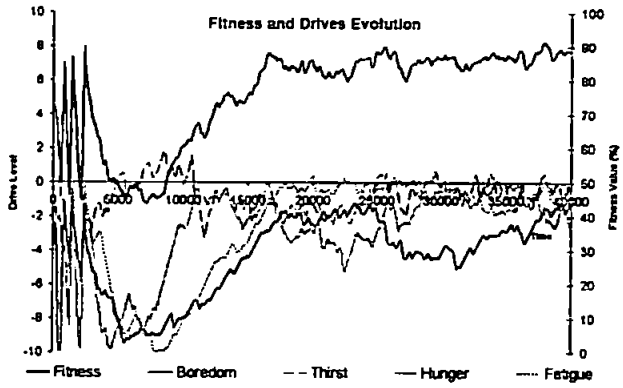


Fig. 7. Fitness vs Drives Evolution.

We propose that these emotional states affect music performance, reflecting the agent emotional state in the music. There are a few studies regarding the communication of emotions through music; for further details please refer to [58]. We simulated different musical performance scenarios inspired by these studies, and the next section presents the outcome from running the Emotional part of our model.

D. Experiment and Result

The objective of this experiment was to analyze the ability of an agent to regulate its homeostasis. To achieve this task we studied the emergence of associations between world stimuli with internal needs; in other words, an implicit world/body map. Fig. 7 shows the relation between fitness function (reflecting the agent's well-being) and the evolution of the agent's drives. The values are averages for each 200 iterations intervals. An overall increase of fitness is shown, suggesting that the agent is capable to adapt itself to new environments. Fig. 7 also shows a decrease of the amplitude of the drives as time evolves. By looking at the evolution of the drives in time we can observe that they were maintained within a certain range. This reflects the ability of the agent to respond to its "body needs". Apparently the agent not only learned how to adapt to the environment, but also did it effectively, maintaining a "healthy behaviour" by self-regulation of the homeostatic process.

A complete analysis of the system is presented in [61] .

E. Performance

Two physiological variables, selected for their influence in actual human performances [58], *Heart Rate* and *Adrenaline*, control *tempo* and *velocity* (loudness) in the performance of a piece of music [62], reflecting neural activity and emotions valence (whether positive or negative), mirroring the agent's emotional state. *Heart Rate* values modulate the on-times of events within each measure (bar), in this case 4000 ms, with a maximum deviation of +/- 640 ms. *Adrenaline* values modulate events' velocity (loudness) between user chosen limits, in this case, 80 and 127. The results can be heard at http://cmr.soc.plymouth.ac.uk/ecoutinho/ (link *Polymnia*).

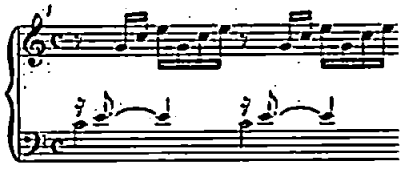


Fig. 8. Score: J.S.Bach - *Prelude no I, BWV 846, from the Well Tempered Klavier I*

Heart Beat	Note duration	Adrenaline	Amplitude	MIDI
71.000	226.387	5.207	92	[145 60 92]
69.263	239.539	9.228	94	[145 64 94]
69.649	246.098	9.652	97	[145 67 97]
69.991	257.054	9.538	97	[145 72 97]
69.770	233.929	9.466	86	[145 76 86]
70.269	264.449	9.962	112	[145 67 112]
70.659	249.279	9.962	98	[145 72 98]
70.548	255.966	10.239	121	[145 76 121]
70.890	237.331	10.335	99	[145 60 99]
70.902	278.871	10.334	102	[145 64 102]
70.908	256.56	10.363	105	[145 67 105]
70.950	250.47	10.288	109	[145 72 109]
71.301	265.184	10.392	85	[145 76 85]
71.323	237.109	10.084	107	[145 67 107]
71.322	261.263	10.078	90	[145 72 90]
71.655	240.51	10.169	112	[145 76 112]

TABLE II
PERFORMANCE DATA (MIDI MESSAGES: [INSTRUMENT PITCH VELOCITY] - PIANO.

We collected the data from the simulation in the previous section to "perform" a piece of music [62]; in this case to playback a MIDI recording of a piece. In Fig. 8 we present the first measure of the piece. The anatomy of each note here represented by three parameters (MIDI messages): note-number, note-duration (measured in ms), and velocity (loudness). In the original MIDI file notes are played every 250 ms. In our piece their duration varies according to *Heart Rate* value(see Fig. II). Velocity (or loudness) is controlled by the level of *Adrenaline*. The system related *Heart Rate* onto music by mirroring stable or unstable situations, relaxation or anxiety with deviations from original rhythmic structure of each measure of music, and *Adrenaline*, by, on the one hand, mirroring excitement, tension, intensity, or, on the other hand, boredom, low arousal, by changes in note-velocity (loudness); refer to Table II.

We are currently testing the model with different conditions and metabolism, specifically the amount of resources needed to satisfy drives and the way in which these drives decrease and increase in time. A deep analysis of the behaviour of the model may reveal that performance in different environments and with different agent metabolisms can play a strong role in the affective states.

V. CONCLUDING REMARKS

At the introduction of this paper we indicated that EA has been used in a number of musical applications, ranging from sound synthesis and composition to computational musicology. An increasing number of musicians have been using EA for artistic purposes since the early 1980s. However, the potential of EA for computational musicology started to be explored only recently, after the works by researchers such as Todd, Kirby and Miranda [23] [24] [25] [63].

This paper presented three components of an A-Life model (using EA) that we are developing to study the development of musical knowledge, rooted on the problem of beat synchronisation, knowledge evolution and emotional systems.

Although the A-Life approach to computational musicology is still incipient, this paper reinforced the notion that a new approach to computational musicology is emerging.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of the Portuguese Foundation for Science and Technology (FCT, Portugal), the Brazilian Ministry of Education (CAPES, Brazil), and The Leverhulme Trust (United Kingdom).

REFERENCES

[1] M. V. Mathews, "The digital computer as a music instrument," *Science*, vol. 142, no. 11, pp. 553–557, 1963.
[2] E. R. Miranda, *Computer Sound Design: Synthesis Techniques and Programming*. Oxford, UK: Focal Press, 2002.
[3] A. Horner, J. Beauchamp, and L. Haken, "Machine tongues XVI: Genetic algorithms and their application to fm matching synthesis," *Computer Music Journal*, vol. 17, no. 4, pp. 17–29, 1993.
[4] R. Garcia, "Growing sound synthesizers using evolutionary methods," in *Proceedings of ALMMA 2002 Workshop on Artificial Models for Musical Applications*, E. Bilotta, E. R. Miranda, P. Pantano, and P. M. Todd, Eds. Cosenza, Italy: Editoriale Bios, 2002, pp. 99–107.
[5] D. Gabor, "Acoustical quanta and the theory of hearing," *Nature*, no. 1044, pp. 591–594, 1947.
[6] P. Thomson, "Atoms and errors: towards a history and aesthetics of microsound," *Organised Sound*, vol. 9, no. 2, pp. 207–218, 2004.
[7] E. R. Miranda, *Composing Music with Computers*. Oxford, UK: Focal Press, 2001.
[8] J. Mandelis and P. Husbands, "Musical interaction with artificial life forms: Sound synthesis and performance mappings," *Contemporary Music Review*, vol. 22, no. 3, pp. 69–77, 2003.
[9] C. Dodge and T. Jerse, *Computer Music*. London, UK: Schimer Books, 1985.
[10] D. Cope, *Computers and Musical Style*. Oxford, UK: Oxford University Press, 1991.
[11] D. Worral, "Studies in metamusical methods for sound image and composition," *Organised Sound*, vol. 1, no. 3, pp. 183–194, 2001.
[12] I. Xenakis, *Formalized Music: Thought and Mathematics in Composition*. Bloomington (IN), USA: Indiana University Press, 1971.
[13] D. Millen, "Cellular automata music," in *International Computer Music Conference ICMC90*, S. Arnold and D. Hair, Eds. San Francisco (CA), USA: ICMA, 1990, pp. 314–316.
[14] A. Hunt, R. Kirk, and R. Orton, "Musical applications of a cellular automata workstation," in *International Computer Music Conference ICMC91*. San Francisco (CA), USA: ICMA, 1991, pp. 165–168.
[15] E. R. Miranda, "Cellular automata music: An interdisciplinary music project," *Interface (Journal of New Music Research)*, vol. 22, no. 1, pp. 03–21, 1993.
[16] B. Degazio, "La evolucion de los organismos musicales," in *Musica y nuevas tecnologias: Perspectivas para el siglo XXI*, E. R. Miranda, Ed. Barcelona, Spain: L'Angelot, 1999, pp. 137–148.
[17] R. Waschka II, "Avoiding the fitness bottleneck: Using genetic algorithms to compose orchestral music," in *International Computer Music Conference ICMC99*. San Francisco (CA), USA: ICMA, 1999, pp. 201–203.
[18] K. McAlpine, E. R. Miranda, and S. Hogar, "Composing music with algorithms: A case study system," *Computer Music Journal*, vol. 223, no. 2, pp. 19–30, 1999.

[19] P. Dahlstedt and M. G. Nordhal, "Living melodies: Coevolution of sonic communication," *Leonardo*, vol. 34, no. 2, pp. 243–248, 2001.
[20] J. Mandelis, "Genophone: An evolutionary approach to sound synthesis and performance," in *Proceedings of ALMMA 2001 Workshop on Artificial Models for Musical Applications*, E. Bilotta, E. R. Miranda, P. Pantano, and P. M. Todd, Eds. Cosenza, Italy: Editoriale Bios, 2001, pp. 37–50.
[21] J. Manzolli, A. Moroni, F. von Zuben, and R. Gudwin, "An evolutionary approach applied to algorithmic composition," in *VI Brazilian Symposium on Computer Music*, E. R. Miranda and G. L. Ramalho, Eds. Rio de Janeiro, Brazil: EntreLugar, 1999, pp. 201–210.
[22] J. Impett, "Interaction, simulation and invention: a model for interactive music," in *Proceedings of ALMMA 2001 Workshop on Artificial Models for Musical Applications*, E. Bilotta, E. R. Miranda, P. Pantano, and P. M. Todd, Eds. Cosenza, Italy: Editoriale Bios, 2001, pp. 108–119.
[23] P. M. Todd and G. Werner, "Frankensteinian methods for evolutionary music composition," in *Musical networks: Parallel distributed perception and performance*, N. Griffith and P. M. Todd, Eds. Cambridge (MA), USA: MIT Press/Bradford Books, 1999, pp. 313–339.
[24] E. R. Miranda, "At the crossroads of evolutionary computation and music: Self-programming synthesizers, swarm orchestras and the origins of melody," *Evolutionary Computation*, vol. 12, no. 2, pp. 137–158, 2004.
[25] E. R. Miranda, S. Kirby, and P. Todd, "On computational models of the evolution of music: From the origins of musical taste to the emergence of grammars," *Contemporary Music Review*, vol. 22, no. 3, pp. 91–111, 2003.
[26] S. Handel, *Listening: An Introduction to the Perception of Auditory Events*. Cambridge (MA), USA: The MIT Press, 1989.
[27] P. Fraisse, *The Psychology of Time*. New York, USA: Harper & Row, 1963.
[28] E. W. Large and J. F. Kolen, "Resonance and the perception of musical meter," *Connection Science*, vol. 6, pp. 177–208, 1994.
[29] E. Scheirer, "Tempo and beat analysis of acoustic musical signals," *Journal of the Acoustical Society of America*, vol. 103, no. 1, pp. 588–601, 1998. [Online]. Available: http://web.media.mit.edu/ eds/beat.pdf
[30] R. Dawkins, *The Blind Watchmaker*. London, UK: Penguin Books, 1991.
[31] A. Cox, "The mimetic hypothesis and embodied musical meaning," *MusicæScientiæ*, vol. 2, pp. 195–212, 2001.
[32] L. M. Gabora, "The origin and evolution of culture and creativity," *Journal of Memetics - Evolutionary Models of Information Transmission*, vol. 1, no. 1, pp. 1–28, 1997.
[33] S. Jan, "Replicating sonorities: towards a memetics of music," *Journal of Memetics - Evolutionary Models of Information Transmission*, vol. 4, no. 1, 2000.
[34] M. Gimenes, E. R. Miranda, and C. Johnson, "Towards an intelligent rhythmic generator based on given examples: a memetic approach," in *Digital Music Research Network Summer Conference 2005*. Glasgow, UK: The University of Glasgow, 2005, pp. 41–46.
[35] S. Strogratz and I. Stewart, "Coupled oscilators and biological synchronization," *Scientific American*, vol. 26, pp. 68–74, 1993.
[36] J. A. Sloboda, *The Musical Mind: The Cognitive Psychology of Music*. Oxford, UK: Clarendon Press, 1985.
[37] "Paderewski." http://www.wyastone.co.uk/nrl/gpiano/8816c.html, Last visited 27 April 2005.
[38] P. White, *Basic Midi*. London, UK: Sanctuary Publishing, 2000.
[39] "An essay on chopin," http://www.geocities.com/Vienna/3495/essay1.html, Last visited 27 April 2005.
[40] C. Darwin, *The Expression of the Emotions In Man and Animals*, P. Ekman, Ed. New York, USA: Oxford University Press, 1998.
[41] W. James, "What is an emotion?" *Mind*, vol. 9, pp. 188–205, 1884.
[42] W. Wundt, *Outlines of Psychology*. London, UK: Wilhelm Engelmann, 1897.
[43] R. Lazarus, *Emotion and Adaptation*. New York, USA: Oxford University Press, 1991.
[44] W. Cannon, *Bodily Changes in Pain, Hunger, Fear and Rage*. New York, USA: Appleton, 1929.
[45] S. S. Tomkins, "The positive affects," *Affect, imagery, consciousness*, vol. 1, 1962.
[46] ——, "The role of facial response in the experience of emotion," *Journal of Personality and Social Psychology*, vol. 40, pp. 351–357, 1981.
[47] R. Plutchik, "Emotion: Theory, research, and experience," *Theories of emotion*, vol. 1, pp. 3–33, 1980.
[48] ——, *The Emotions*. University Press of America, 1991.
[49] C. E. Izard, *The face of emotion*. New York, USA: Appleton-Century-Crofts, 1971.
[50] ——, *Human Emotions*. New York, USA: Plenum Press, 1977.
[51] ——, "Differential emotions theory and the facial feedback hypothesis activation," *Journal of Personality and Social Psychology*, vol. 40, pp. 350–354, 1981.
[52] ——, "Four systems for emotion activation: Cognitive and noncognitive processes," *Psychological Review*, vol. 100, no. 1, pp. 68–90, 1993.
[53] P. Ekman, "Basic emotions," in *The Handbook of Cognition and Emotion*, T. Dalgleish and T. Power, Eds. Sussex, UK: John Wiley and Sons, Ltd., 1999, pp. 45–60.
[54] ——, *Darwin and facial expression: A century of research in review*. New York, USA: Academic, 1973.
[55] A. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, USA: Avon Books, 1994.
[56] ——, *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. London, UK: Vintage, 2000.
[57] ——, *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*. New York, USA: Harcourt, 2003.
[58] P. N. Juslin and J. A. Sloboda, Eds., *Music and Emotion: Theory and Research*. Oxford University Press, 2001.
[59] E. Coutinho, H. Pereira, A. Carvalho, and A. Rosa, "Livia - life, interaction, virtuality, intelligence, art," Faculty of Engineering - University of Porto (Portugal), Tech. Rep., 2003, ecological Simulator Of Life.
[60] N. Gracias, H. Pereira, J. A. Lima, and A. Rosa, "Gaia: An artificial life environment for ecological systems simulation," in *Artificial Life V: Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems*, C. Langton and T. Shimohara, Eds. MIT Press, 1996, pp. 124–134.
[61] E. Coutinho, E. R. Miranda, and A. Cangelosi, "Towards a model for embodied emotions," *Accepted to the Affective Computing Workshop at EPIA 2005*, 2005.
[62] J. S. Bach, *The Well-Tempered Clavier, Book I*, vol. BWV 846, ch. Prelude I (C major).
[63] E. R. Miranda, "Creative evolutionary systems," in *On the Origins and Evolution of Music in Virtual Worlds*, P. J. Bentley and D. W. Corne, Eds. Morgan Kaufmann, 2002, ch. 6, pp. 189–203.

Coutinho, E., Miranda, E., & Cangelosi, A. (2005). Artificial emotion - simulating affective behaviour. *In Proceedings of the Post-cognitivist Psychology Conference.* Glasgow, Scotland.

Artificial Emotion: Simulating Affective Behaviour

Eduardo Coutinho - Eduardo Reck Miranda - Angelo Cangelosi

University of Plymouth
Plymouth PL4 8AA
{eduardo.coutinho, eduardo.miranda, angelo.cangelosi}@plymouth.ac.uk

Artificial Life (A-Life) is new area of research within Artificial Intelligence and Cognitive Sciences which focuses on the role of bottom-up processes in the self-organisation of behavioural and cognitive processes.
This paper describes an A-Life simulation of emotional behaviour in autonomous agents. It aims to model part of the emotional process where self-survival tasks and embodiment influence behaviour and affective states. The model consists of a population of agents with complex cognitive, emotional, and behavioural abilities. Agents live in an environment where they can interact with several objects related to their behavioural and emotional states.
The agents' cognitive systems can be described as consisting of three main parts. (1) The Perceptual System receives information from the environment through a retina. (2) The Behavioural System, organised in a Motivational and a Motor Control sub-systems, which define the agents' interaction with its environment. (3) The Emotional System, which considers the role of emotions as part of the homeostatic mechanisms.
An agent internal body state is defined by a set of physiological variables that vary accordingly to their interaction with the world, and a set of internal drives. The agent is controlled by a feed-forward neural network that integrates visual input and information on its internal and emotional states to interact with objects and explore the world. The network learns through a reinforcement learning algorithm.
Simulations aim the analysis of the agent ability to regulate its homeostasis. Results show that the agent manage to organize its behaviour towards an "*equilibrium*" state (homeostatic regime). Specifically, when simulating different drives variations (different body needs), we observed a bias for the agents' motivations, reflecting drives urgency, cycles, and emotional state.
Current simulations are focusing on modelling higher level emotional structures. As an example, we are currently looking at differences between different emotional and motivational processes, such as temperaments and complex drives (e.g. social drives), as well the influence of music in emotional states.

Keywords: emotions, neurobiological modelling, artificial life, reinforcement learning, affective computing, music.

3524-1