INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

I-Sounds

Emotion-based Music Composition for Virtual Environments

Ricardo Miguel Moreira da Cruz

Dissertação para obtenção do Grau de Mestre em

Engenharia Informática e de Computadores

Júri

Presidente:	Prof. Ernesto José Marques Morgado
Orientador:	Profª Ana Maria Severino Almeida e Paiva
Vogais:	Prof. Fernando Amílcar Bandeira Cardoso
	Prof. Eduardo Lopes

Abril 2008

Acknowledgements

There are certain moments in life, that we stop and think; about the past, the brief present and the future of all hopes. This is one of those moments. Soon, I will be flying my own life, pursuing my dreams.

The first persons that I would like to thank, are my parents and brother. To my parents for their hard work, continuous support and always comforting love. After all these years of study I have always looked to correspond at the best of my efforts, to the hopes and aspirations put on me. To my brother, the most important person in my life, I would like to thank his full-time commitment in my education. He has most of the responsibility of what I am. All this academic path would be extremely hard, without his continuous support, stimulus and motivating attitude. With him I learnt to ask "Why?". A big thank you to my family, I will be in eternal debt with you.

This work is the product of eighteen months of hard work and extreme dedication. The road, would have been a dark and cold place to travel, without the lights of my supervisor, Prof. Ana Paiva, co-supervisor, Prof. Eduardo Lopes, and the important role of my proximity supervisor and friend, António Brisson. I would like to thank, Prof. Ana Paiva, for her willingness, support and the work conditions that she generously provided, Prof. Eduardo Lopes, for his commitment in my work in spite of his professional activity in other university, and for all the important meetings that he was always willing to have. Engineering and art, can definitely meet for extraordinary outcomes. To António, my eternal gratitude for the last year's work, his continuous support and motivation "snippets", I-Sounds would not have reached what it is today without your cooperation. The best wishes for your starting career.

To all the people at IST Tagus, thank you for your friendship. To all my friends, which are so many that it would be hard to squeeze them here in this page. How I feel lucky for that! You made these last 5 years an incredible experience. There are however, two of you, who deserve a special place. To Diogo and Sara, my eternal partners and friends, always there in the bad and good moments a sincere thank you. I remember now, the days and nights spent in front of the computer, the ideological fights with you Diogo, the stubbornness of the talks with you Sara, the dinners and the nights spent at Bairro Alto, studying for the exams... Your invaluable brotherhood is the most rewarding gift of all these years. I am a very lucky guy, in having both of you at my side, thank you...

I would also like to thank all the people involved in the GAIPS project. Our vision of today, will shape the future of how we see the machine. It was an honour to work in the forefront of this exciting and new world, with this research team. A word of gratitude to Carolina, Leila and Sara my office partners. Two special thanks to Guilherme Raimundo and Rui Prada, for their statistical advisory. My decision to follow other path would not be a cut with the place where it all started. A big sincere, and "affective", hug for all of you in GAIPS.

Thanks to all of you!

Ricardo

Abstract

English

From the concert halls to our pockets, music has an ubiquitous presence in our lives. Due to its observable, yet not fully understood, expressive abilities, it is a very effective communication tool explored in computer games, films and virtual and interactive systems. In spite of the crescent demand for flexibility, nowadays, most applications use pre-composed soundtracks, less flexible though aesthetically refined.

A new trend takes the paradigm a step further to develop automatic composition systems, able to deliver real-time contextualised music, while the user is presented with an every changing interactive experience. This work is above all an exploratory proof of concept and a step towards that objective, proposing a computational architecture and implementation for an emotion-based composition system. The goal is to provide a composition algorithm development framework, as well as, a run-time environment for integration with affective systems. In parallel, this work proposes a composition algorithm able to express happiness, sadness, anger and fear, using the properties of rhythm and diatonic modes.

In conjunction both systems are expected to enlarge the affective bandwidth of an Interactive-Drama application, in which children can build up stories with the collaboration and participation of computer controlled characters. While the collected empirical data does in fact supports the hypothesis, it also uncovers some aspects requiring further refinement, such as, the illustration of anger and fear. The experiments, have also validated the I-Sounds framework as a suitable development and integration tool for composition algorithms.

Keywords: Affective Computing, Emotion, Music, Automatic Composition, Rhythm, Mode

Portuguese

Das salas de concerto até aos modernos leitores de bolso, o carácter ubíquo da música é uma realidade. A sua expressividade óbvia, contudo ainda largamente incompreendida, constituem uma ferramenta de comunicação poderosa explorada em vídeo-jogos, filmes, ambientes virtuais e interactivos. Apesar da crescente exigência por uma maior flexibilidade a maioria das aplicações, de hoje continua a usar

bandas sonoras pré-compostas, musicalmente sofisticadas mas pouco flexíveis.

Em resposta, uma nova vaga de investigação pretende revolucionar este paradigma desenvolvendo sistemas de composição automáticos capazes de compor música contextualizada em tempo real, melhorando assim a interacção com o utilizador. Este trabalho é sobretudo uma prova de conceito e um passo naquela direcção, propondo uma arquitectura e implementação para um sistema de composição baseado em emoções, i.e., uma “framework” e ambiente de execução para o desenvolvimento de algoritmos de composição, integrável com qualquer sistema afectivo. Ainda no âmbito deste trabalho é desenvolvido um algoritmo para a expressão de felicidade, tristeza, raiva e medo, usando para isso as propriedades do ritmo e dos modos diatónicos.

É esperado que os dois sistemas melhorem a comunicação sistema-utilizador, num ambiente interactivo para geração de narrativas, através do qual as crianças podem construir histórias, colaborando com personagens controladas por computador. Os dados experimentais recolhidos são encorajadores e apontam claramente nesse sentido, contudo, revelam pontos de possível aperfeiçoamento, tais como, a ilustração musical da raiva e do medo. Os resultados validam também a "framework" proposta como ferramenta adequada ao desenvolvimento e integração de algoritmos de composição.

Palavras chave: Computação Affectiva, Emoções, Música, Composição Automática, Ritmo, Modo

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Notes are less important in music, but the sensations elicited by them."

Leonid Pervomaisky

Music is an art form consisting of sound and silence and one of the most extraordinary manifestations of the human intellect creativity. Music is believed to have appeared among the first pre-historic human communities serving both entertainment and practical purposes, such as, communication and luring of animals during hunt. Music has been a subject of great interest for Science and Philosophy, many scientific disciplines address music from a specific point of view, Psychology, Anthropology, Biology and of course Musicology are in the among the most important. These different perspectives share a common goal, to understand the phenomena - natural and psychological - involved in the composition, performance and listening processes. A subset of these studies are focused on musical emotions, i.e, the relation between music and emotions. Indeed, the affective dimension of music is at the same time one of the most evident - no one would deny it - and difficult to study. Today music has an ubiquitous existence on virtually all human activities being a very important part of our collective culture. The computing area is not an exception, and nowadays with the technological advances it is possible to compose, record, encode, distribute, decode and play music with the help of the computer. In fact digital recording, processing and distribution of music as long surpassed the analogue paradigm. This work aims to provide new research tools and applications, exploring the frontier between music and our emotions

1.1 Motivation

Soundtracks and sound effects are an essential part of games, films and other entertainment applications. In such contexts music serves a broad set of functions, such as emphasising something relevant to the player or spectator, create suspense, or induce emotional reactions. However, the ultimate goal is almost always the improvement of the user's or audience's experience while enjoying the interaction with such applications.

On most applications, soundtracks are previously composed and integrated into the game or film. This approach can have different refinement levels but they general offer the same advantages; first, the music is generally created by artists whose creativity and artistic vision entail an high aesthetic value, which is transmitted to the audience, and secondly, because everything is composed *a priori*, the artist is able to do a careful study of the "script" and experiment various alternatives until he is satisfied, enabling better musical "illustrations". However, this method fails to provide the increasingly demanded flexibility, e.g., changes to the original script may enforce changes to the soundtrack which are time and resource consuming, and of course the user looses the interest quickly while having a less immersive experience. Computer games are an example in which a higher level of flexibility in the course of action, the possibility to instantly adequate music and sound effects to the action and the ability to give the user a slightly different soundtrack every time he plays, would contribute to a better gaming experience. Would it not be nice to customise a game's soundtrack to fit our personal taste, and not be constrained to a list of available options which incidentally we do not particularly like, ending up to turn the music off? This issue implies the control over the processes of composition and performance. Although automatic or semi-automatic music composition lacks the aesthetics of a real composer it is the best option to provide the desired flexibility.

Such as most works the original idea and concept for this thesis born from previous research, more precisely from an interactive system which felt the need for flexible contextualised music. The I-Shadows project [4] proposes an Interactive-Drama application, which provides an affective environment where children can build up stories with the collaboration and participation of computer controlled characters - affective agents. Among the numerous challenges presented by I-Shadows there is the bidirectional communication between system and users. The system proposes the usage of the affective paradigm, i.e., emotions, as the common system and user language. A crucial aspect of this approach is that to cooperatively build a story, both the computer and users should be able to recognise the emotions exhibited by the characters controlled by each other. The recognition of the characters emotional state is based on the motion patterns they exhibit and restricted to the visual channel. The users naturally observe the projected characters and perceive their emotions while the computer uses a video camera to capture and interpret the collected plot images. Good results have been achieved, particularly in the user-system direction, and although the ones achieved in the opposite direction are also encouraging, there is some space left for improvement. The usage of an additional communication channel, the auditive, is expected to enable such improvement.

1.2 Objective

Starting from this context, the present work aims to develop a general solution that can provide not only I-Shadows but a variety of other affective systems with a flexible auditive channel, i.e., a flexible solution that can integrate with existent affective systems and enable a richer interaction. Therefore this thesis can be divided in three main parts: the architectural proposal for a composition algorithm development framework, as well as, an emotion-based run-time environment to integrate those algorithms with any affective system, a full implementation of the proposed architecture, and finally, the development of a concrete affective composition algorithm. These components fulfil the following research objectives;

1. The definition of suitable development tools for emotion-based composition algorithms.
2. The usage of music theory in the development of a concrete composition algorithm.
3. The improvement of an interactive system's users experience and immersion, through the usage of contextualised music.

It is adequate and somehow mandatory at this point, to advance the hypothesis that this work evaluated and aims to prove. A similar enumeration can therefore be made;

1. The proposed emotion-based composition system is a suitable tool for the development and integration of composition algorithms.
2. Effective composition algorithms can be built upon the formalisation of theories addressing music's structural factors and their perceptive qualities.
3. Music can be used to improve the user's interactive experience and overall intelligibility of interactive systems.

1.3 Outline

The remaining of this document is divided into six chapters. A theoretical and empirical framework for the present work is presented in the literature review of chapter 2. It starts with a brief discussion of conceptual models of emotion with a special emphasis on the bi-dimensional approach. The chapter proceeds with a multidisciplinary perspective, especially Musicological and Psychological, over music and emotion, followed by a comprehensive anthology of relevant empirical studies on musical emotions. After this review and reflecting the prominent role of rhythm in this work, an entire section is dedicated to the exposition of the theoretic foundations used in the development of a composition algorithm. Once the relevant topics on emotion and music are covered the field of computer music and relevant practical applications receive some attention in a concise review organised after different approaches to the topic. The chapter finishes with an interesting case-study on the usage of music to improve the audience's experience and immersion in the film industry.

Chapter 3 proposes a layered architecture for a composition algorithm development framework and emotion-based run-time environment. The chapter begins with the description of some high-level requirements and continues with the architectural specification.

Chapter 4 discusses the high-level options of a complete implementation for the proposed architecture.

Chapter 5 discourses on the implementation of a rhythm-based composition algorithm and its auxiliary melodic elements. The chapter discusses the algorithm's theoretical foundations in detail and proposes a formalisation suitable for a computational implementation, and proceeds with the discussion of how rhythm and the diatonic modes are related with the expression of emotions. Finally the chapter discusses some relevant implementation aspects of the composition algorithm.

Chapter 6 presents the three experiments carried out for this work. The first experiment evaluates the composition algorithm in a purely auditive environment and provides calibration data to parametrise the algorithm. The second experiment measures the effect of music on the recognition of emotions exhibited by animated characters. Finally, the third experiment measures the qualitative improvement of the users interactive experience and story intelligibility under an Interactive-Drama application.

The contribution and future developments of this work are the matter of chapter 7, which also provides a small annotation about the field of musical emotions and their practical applications, from which this work is an example.

Chapter 2

Literature Review

"Every music that paints nothing, is nothing but noise."

Jean Alembert

Music as the remarkable ability to make us smile and cry as no other art or activity. It is common to use adjectives semantically related with affects to classify and describe music. This fact is valid both to listeners and musicians whose artistic and aesthetic ideas are better expressed that way. Not only music influences our emotional state, but our emotional state can also influence our musical choices of the moment. The relation between music and emotion is therefore a difficult topic to comprehend. Dowling and Harwood stated [18];

> "...music arouses strong emotions in people, and they want to know why..."

Surprisingly - or not so surprisingly considering its difficulty - it was not until the beginning of the current century that the study of musical emotions left the list of neglected topics. Two of the most important and influential books on music psychology [17] and emotion psychology [54] do not contain an in-depth analysis of musical emotions. Some reasons for this neglect such as, the problem of studying affective phenomena in laboratory, are shared with the emotion field, but others are specific to the topic of musical emotions. Benefiting from an unifying paradigm - in contrast with musical emotions - the prevalence of Cognitive Science put the emphasis on cognitive aspects of musical behaviour, both in perception and performance, additionally the lack of theories, the non trivial description of emotional responses to music and the emergence of Behaviourism - 40's and 50's – lead to problem avoidance and abandoned research. In its turn early studies before 1940 were more concerned with matching affective labels and music excerpts, lacking a general sense of direction. Kate Hevner [32, 33] pioneer studies are the exception confirming the rule, and her work still influences modern research on musical emotions in a pragmatic manner.

This state of affairs changed radically, with the beginning of the new century. In 2001, Patrik

Juslin and John Sloboda edited the first scientific anthology [39] exclusively dedicated to the field of music and emotion. The book provided the much needed and waited referential ground that could encourage and support the renewed interest on music and emotion, an undeniable reality in the research community. An increasing number of articles mentioning music and emotions, now appear in conference proceedings. The present work is strongly influenced by that important contribution and it is part of this new wave of exciting and innovative research on the topic.

2.1 A Dimensional Look of Emotion

Questions about the nature of emotions are as old as human thought, yet it seems there are no definite answers. Interesting enough is the fact that in spite of being an essential part of human intellect, with its dynamics apparently understood by each of us, a lack of formal understanding remains. Psychologists study emotion for almost half a century – since the 19th century - but research lights were for the most part turned into other problems and higher mental processes, such as reasoning, problem solving and decision making. One of the psychological questions concerning emotion is its conceptualisation. A number of different approaches exist, and although this section focuses on dimensional approaches, two other are briefly mentioned due to its relevance and complementarity.

The categorical approach focuses on the characteristics that distinguish emotions among them. Central to these approaches is the idea of a limited number of innate and universal emotion categories, represented by basic emotions, in which all other emotions can be classified. Categorical approaches have been criticised on the base argument of no consensus between researchers on the set of basic emotions; although a general agreement on five of those emotions: happiness, sadness, anger, fear and disgust.

Dimensional approaches were among the first to appear. The common idea behind these approaches is the mapping of emotions into a dimensional space, defined by two or more dimensions, such as valence, activity and potency, in which similarity relations are established based on dimensional proximity. Categories can also be derived from a dimensional space; however, they differ from the categorical concept in that they are based on a reduced set of variables, i.e., the model dimensions.

One of the most notable dimensional approach is Russel's *circumplex model* of emotion [79], consisting of two dimensions, *valence* (horizontal) and *activation* (vertical). Emotions are mapped into the bi-dimensional space forming a circle. Two emotions are said to correlate inversely if they are found across the circle, e.g., happiness and sadness. Russel's model captures two important aspects of emotions, different levels of similarity and the bipolar relation between some emotions, e.g., happiness and sadness. The horizontal dimension of Valence provides a measure of affective appraisal while Activation measures physiological activity. The model's ability to classify each emotion in terms of its appraisal and activity is seen as one of its strengths. Dimensional models also receive a dose of criticism, based on the argument that their dimensional simplification may ignore or have a softening effect on some important psychological differences that can exist between two emotions closely rep-

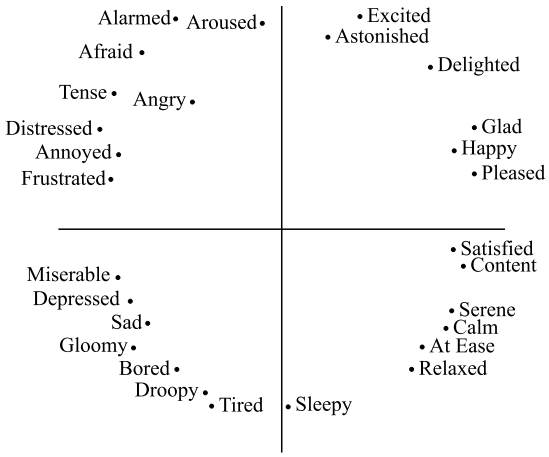


Figure 2.1: Russel's circumplex model of emotion

resented, e.g., anger and fear in the circumplex model. In spite of their differences categorical and dimensional approaches try to capture different aspects of the same thing, and thus they are often seen as complementary rather than alternatives.

The third approach is somehow a compromise between the two previous approaches. The prototype approach not only addresses each emotional category's characteristics but also its hierarchical relationships. A central concept is that of, prototype, an abstract image whose features are representative of a family. Each emotion associates with a prototype according to the resemblance level of relevant features, such as elicitors, cognitive appraisals and physiological reactions. Prototypes do have a natural resemblance with the concept of basic emotions from the categorical approach, and proximity relations, a characteristic of the dimensional models, can also be established through an hierarchy of prototyped groups.

The scope of this work follows a dimensional approach inspired in Russel's circumplex model of emotion. There is evidence of the same circular topology in music [89] and these models have been used successfully to capture and represent continuous change of emotional expression in a comprehensive way. Emery Schubert [85] proposes an extensive study on the continuous measurement of emotional expression using a bi-dimensional model. The author also contributes an interesting chapter [86] on the same topic.

2.2 Music and Emotion a Multidisciplinary Perspective

The dynamics of musical emotions are primarily a psychological problem, and Psychology is the better equipped discipline to study the mental processes evolved in music perception and/or induction of emotions. However, it is far from being alone and much can be contributed by other disciplines. This same idea is sustained by Juslin's and Sloboda's book [39], upon which this section is based.

Philosophy is the first to contribute with its ability to deal with conceptual problems, very common in the music and emotion field and requiring urgent attention. In fact Philosophy was the first

discipline to address these questions. The first music treaties can be dated back to ancient Greece by the thoughts of Plato and his student Aristotle. By that time it was already acknowledge that music has the ability to influence our emotional state, and references to affects can be found on those ancient treaties. Three essential problems are considered by Philosophy. First music is perceived as expressive of emotion, i.e., it is able to be happy and sad itself, although it is not a sentient being. Often the listener's emotional response mirrors the music, and although this phenomena usually lie in the existence of beliefs, there is no evidence of those supporting beliefs. Finally listeners seem to enjoy revisiting musical pieces that makes them feel sad, which is an infrequent pattern of behaviour. Davies [15] discusses major philosophical theories that addressing the above questions, and proposes a set of constraints that should be respected by theoretic formulations of music expressiveness.

Musicological studies on the topic can also be traced back some centuries ago. In fact Juslin and Sloboda [39] sustain that musicological and psychological approaches are only separable for the last three centuries. Cook and Dibben contribute a musicological point of view [9], centred in theoretical frameworks that can explain music's ability to express, induce or even embody emotional states. There is a particular and interesting conclusion emerging from Cook's and Dibben's analysis; the authors sustain that the understanding of musical emotions is strongly related with the contemporary conceptions of emotion itself, i.e., the historical context influences the way Musicology, as well as other disciplines, consider musical emotions. Cook and Dibben proceed to review some classic musicological theories concerning the topic of musical emotions in which two approaches are given special attention, the Neo-semiotoc approach of Hatten's, pointed as the most suitable and coherent approach to integrate expressive and structural analysis, and the performative approach, that analyses music in terms of how "it is herd" and not of how "it is". The authors conclude that contemporary musicological knowledge influences our contemporary perception of music, suggesting that the simple act of speculating about the emotional content of a particular music, may influence us to perceive it as expressive of those emotions.

For psychologists the central questions is to understand the mental processes of music perception, i.e., what happens between our ears and the perception of an emotion. Juslin and Sloboda [40] consider two primary sources of emotion in music, intrinsic and extrinsic. Intrinsic sources are closely related with certain music structural characteristics. Constructions, such as syncopation, enharmonic changes and melodic appogiaturas are closely related to the management, creation, maintenance, confirmation and disruption of musical expectations. Extrinsic sources can further divide into iconic sources emerging from natural similarity between music structure and some event or "agent" suggesting a particular emotion and associative sources emerging from casual associations between music and other external factors carrying some type of affective information. Our memory is able to associate certain types of stimuli, including musical ones, with specific events and contexts. Such associations seem to be stronger when they were the consequence of emotionally strong events [21]. Some investigations found certain music pieces to trigger association mechanisms, and usually when that occurs our mind changes its focus to the original event, i.e., the one that once triggered the now evoked association.

The eliciting of high energy emotions such as, excitement and anger, by high energy loud and fast music is an example of an extrinsic iconic source. To explain extrinsic associative sources, Juslin and Sloboda quote a famous example given by Davies [14], to whom associative sources are of the type;

"Darling, they're playing our tune"

Biology is also concerned with musical emotions. Particularly biologists are concerned with the cognitive organization of musical experience. Brain studies are expected to shed some light over the cognitive processes involved in music perception. Biology addresses the question from different perspectives: physiological, evolutionary and neuropshycological - the later discussed by Peretz [69]. All of them take advantage of the recent and exciting advancements on Brain-Imaging Techniques, such as Positron Emission Tomography (PET), Functional Magnetic Resonance Imaging (fMRI) and Event-related Potentials (ERPs). These front-edge techniques provide a huge flow of data that can be used to "map" musical emotions and its associated processes into brain regions.

Anthropologists are interested in how musical emotions manifest and evolve in different cultural contexts. Becker [2] presents some perspectives on how music is made and how music is used as an instrument of cultural expression. Particularly interesting is the concept of "habitus" of listening, introduced by Becker and used to describe different habits of musical listening and listeners expectations across different cultures. These habits influence the relations of listeners and music itself. They imply assumptions about music meaning, its function, how it should be perceived and suitable reactions to it. Becker points out that affects, are also culturally susceptible, in its conception, expression and perception.

Naturally, Sociology has the same "social" attitude towards music and emotion - eventually more accentuated. The traditional focus on music production changed to music consumption and music's actual social role. DeNora [16] observes that music can be conceptualised as a tool for emotive action. In the author's opinion a promising trend is the study of how music can help actors to create and maintain "characters" identities, embodying their emotions. Recent studies by Hennion and DeNora show that listeners are rather active when affected by music, changing their ability to be influenced. Perhaps more interesting is the evidence on how people use music to actively regulate, enhance and change their emotions. Could we possible be aware of the music we need now and then? DeNora's studies show that this awareness is directly related with associations between certain musical materials and extra musical events. It is opportune at this point to recall the extrinsic associative sources of psychologists, that seem to support this hypothesis, but DeNora notes another interesting parallelism; the sociological associative concerns are in accordance with musicological questions about how certain music structures are more susceptible to particular interpretations.

2.3 Musical Structure and Emotional Expression

The relation between specific factors of musical structure and expressed emotions has long occupied thinkers and scientists, as well as artists; however, the empirical study of musical expression is much more recent, starting in the late nineteenth century. This section intends to be a small anthology of such studies, especially after 1930, but three clarifications should be made at this point.

It is important to distinguish between two different phenomena, the perception of emotion, and the emotional reaction to music. There is a difference between classifying a music as being happy and to feel happy while listening it. This discussion is closely related with the distinction between cognitivism and emotivism - suggested by philosophers, such as Kivy [44]. Cognitivists defend that music can express emotions, that in its turn, can be perceived by the listener, while emotivists defend that music has the ability to elicit real emotional responses in listeners. Scherer and Zentner [83], comment the following on this;

> Our view is that it would be premature to prejudge the issue and that both positions may be perfectly appropriate depending on a number of factors.

This work shares the same view; however, the empirical review in section 2.3.1 starts from the assumption that music is expresses emotion. Thus, the present work focuses on perceived emotions, and so the remaining of this section, with exceptions being conveniently annotated. The alternative approach, is considered in the contributed chapters of Gabrielsson [26], Scherer and Zentner [83] and Sloboda and O'Neill [87].

Music's expressiveness results from the composition and performance processes, thus, it is important to distinguish the properties of the composed music and of its performance. The separation is rather thin; however, the definition of music performance should help to clarify its role on emotional expression. Music performance can be defined as being the result of the slight deviations a performer does while playing the nominal values notated in the score. Those deviations are more common in tempo and dynamics (sound intensity). Sometimes the composer provides special directions in the score, encouraging the performer to do such deviations for the benefit of piece's expressiveness. The performer can effectively, through the slight manipulation of tempo and dynamics, imply a feeling emphasising that side of the piece. In many situations, such freedom is related with style, personal interpretation, emotional state or aesthetic considerations. Thus, music performance is complex and a very personal activity involving physical, acoustic, physiological, psychological and social/artistic issues. The present work focuses on composition rather than performance, and thus, all the following discussion is done from a composition point of view.

Finally, some of the studies referenced in section 2.3.1, date from the early and middle 20th century. One can question the scientific relevance of such studies, after all music perception is influenced by the historical context; however, modern research is using the the results and methodologies of such pioneering studies. Most of them remain actual and their importance is sufficient to justify their reference.

2.3.1 Musical factors and their influence on emotional expression

Alf Gabrielsson and Erik Lindström contribute a review [28] of the most relevant studies about musical structure and emotional expression. Before discussing each factor, it is important to say a word about the different methodological approaches that have been used by different studies. The earliest studies used real music, i.e., existent pieces, ensuring good ecological validity, but making the assessment of each factor's individual contribution a hard task, because they were perceived as a whole. To difficult the task even further, researchers were forced to use live performances before technological advances allowed a reasonable recording quality, as a consequence performance properties inevitably influenced the results; however, their effect is very hard to isolate. In these studies different report techniques were used; free phenomenological descriptions, choice among descriptive terms and ratings of how well such terms apply to a specific music, are among the most common. The systematic variation of each individual musical factor using short sequences without musical context, is the alternative that sacrifices ecological validity for more accurate results. Finally, in an attempt to combine the best of the two worlds, other scientists used the systematic variation of one or more factors using real music segments. But this method has its own limitations, the manipulation of certain factors in real music can lead to very unnatural odd music, resulting in limited or complete lost of ecological validity.

The remaining of this section discusses one structural factor at a time, and it is largely based on Gabrielsson's and Lindström's article. The following discussion, inevitably, uses many terms and concepts specific to the musical domain. To help the unfamiliar reader, each of these terms and concepts is defined in the glossary available in annex A.

Tempo

Tempo is found to be one of the most, if not the most important factor of emotional expression. Fast tempo may be associated with expressions of activity/excitement, potency, surprise, fear [82], happiness [33, 97, 45, 62, 98, 25, 81, 38], joy [75, 1] and anger [82, 38]. On the contrary, slow tempo is usually associated with serenity [33, 31], solemnity [75, 98], sadness [33, 97, 45, 3, 98, 81, 82, 38, 90, 50, 70, 1], tenderness[38], boredom and disgust [82].

An interesting observation is that of tempo notation in musical scores. Tempo can be expressed in beats per minute (BPM) indicating the number of pulses (e.g. 100) of a certain type (e.g. Quarter note) per minute. A notation of, "quarter=60", means that each quarter note has the duration of a second. Thus, it is possible to calculate the duration in seconds or fractions of it, of each note or rest, being also possible to calculate the nominal duration of an entire piece. However, sometimes the composer does not provide the exact tempo in bpm, providing a qualitative description in the form of adjectives, instead. Those adjectives usually appear in Italian, although other languages might be found as well; examples include allegro, presto, andante, adagio, largo, etc. All these words already imply a certain affective connotation, but composers felt the need for intermediate expressive levels. To achieve such purpose they use qualifiers, such as assai, non troppo, molto, poco, più, etc.

Additionally, mood marks can also be used to inform about the piece's character, these marks include terms, such as agitato, dolce, maestoso, sostenuto and vivace. The performer's primary tool when it comes to fulfil the composer's annotations is the manipulation of tempo. Thus, not only the empirical results show a strong influence of tempo in emotion expression, but there is also a strong evidence in music terminology, suggesting that this is acknowledge by musicians for a long time.

It is important to say that tempo might have different meaning, this same fact is pointed by Gabrielsson and Lindström [28]. Sometimes, depending on rhythmic and metric organisation, perceived tempo differs from the musical tempo. In what respects to emotional perception the relevant measure is, perceived tempo, a concept explored by the composition algorithm discussed in chapter 5.

Mode

Mode has a strong expressive potential. Albeit the blurring and sometimes confusing definitions for musical mode, persisting in music for a long time, it seems that musicians agree about their functionality, and although they might conceive it differently, they use it in similar ways. The term mode is usually applied in the context of the diatonic scales, therefore modes can be of two types, major and minor. The later may assume different formulas, but that would not be considered here. Whenever minor mode is referred its harmonic formula is assumed.

The major diatonic mode is usually associated with happiness [32, 45, 98, 82, 13, 50, 70] and joy [73] while the minor diatonic mode is usually perceived as sad [32, 45, 98, 13, 50, 70]. Major mode may be defined as graceful, serene, dreamy, dignified [32] and solemn [73], and minor mode associated with tension [81], disgust and anger [82]. However, the perceived emotion is highly dependent on the musical context. e.g., fast tempo and high pitched chords may be perceived as happy regardless of mode, but the general tendency is to perceive major mode as happy and minor mode as sad, at least in western music culture. Musicians are aware of this, and probably one of the best examples that can be pointed out is that of Fado, a traditional song of Lisbon and Portugal. Traditional Fado is usually classified in three categories: "Fado Menor", "Fado Mouraria" and "Fado Corridinho". From these, the first type named after its minor scale, is defined as the most sad of all fados, a bittersweet of melancholy.

Loudness

Loudness is sometimes referred as a parameter of dynamics. In either case it refers to the relative "volume" of musical notes. Some instruments, namely the piano are able to play the same note with different loudness or intensity. Along with tempo this is one of the most important aspects of music performance. The performer usually adjusts loudness levels to fulfil its interpretation and artistic conception of the piece. Composers also provide dynamics indications in the musical score using special symbols, such as ff, f, mf, p, and pp, which are abbreviations of fortissimo, forte, mezzo forte, piano, pianissimo respectively. These indicate the "strength" or emphasis that should be put in each note. Progressions and transitions between different levels can also be notated by the composer,

usually named crescendo or diminuendo. The usage of dynamics marks is rather free among composers and performers. The symbols above are the most common and normal, but exaltations may also be found. Tchaikovsky often considered a very emotional composer indicates a bassoon solo as "pppppp" – very very gentle - in his "Pathétique" symphony and "fffff" in passages of the "1812 Overture". Rachmaninof and Shostakovich made use of such extreme dynamics notation, but an interesting case is that of Gustav Mahler's that instructs the violins to play "fffff" in the second movement of his seventh symphony, along with an explicit written note in the score saying, "pluck so hard that the strings hit the wood".

Loud music is associated with intense emotions [98], power [45], tension [64, 49], anger [38] and joy [73]. Soft music is perceived as soft [45, 98], tender [45, 38], sad, solemn [38] and sometimes as frightening. Loudness variation seems to affect emotional expression as well. Large variations may be associated with fear and small variations with happiness [82]. Timing is also important, rapid changes are associated with a playful mood [97, 50] and few or no changes with sadness, peace, dignity and monotony [97].

Pitch

Pitch level has a natural influence in musical expression. High pitch music is often perceived as happy, serene, dreamy [33] and expressive of surprise, anger and fear [82]. Low pitch is associated with sadness [33, 97, 98, 82], solemnity [74, 98] and boredom but sometimes with pleasantness [82], depending on the overall musical context. Pitch variations may also account for expressiveness, high variations associate with happiness and small variations with anger and fear [82].

Intervals

In music, the term interval usually refers to the difference between two notes. There are two types of intervals, harmonic and melodic. The first are the base of chord construction and refer to the difference between two simultaneous notes. Melodic intervals refer to the difference between two consecutive notes. The following applies only to melodic intervals, their harmonic counterparts are treated in the topic of harmony, below. Empiric evidence on the relation between intervals and emotion perception is rather scarce. However, it seems that larger intervals are stronger than small ones. Minor second is referred as the most sad [57], and the octave as the most positive [88]. Melodic intervals, as well as, their harmonic counterparts are highly cultural dependent, e.g., referring the minor second or the octave might only be appropriate in diatonic contexts and others influenced by western musical culture.

Melody

Melody is usually studied according to its range and direction. Wide melodic ranges are associated with joy [1] or fear [50], and a narrower range with sadness [1], sentimentalism and calmness [31]. Ascending melodies are perceived as serene [32] and happy [29], they are also associated with fear,

surprise and anger [82]. Descending melody is exciting, graceful and vigorous [32], but sometimes sad if combined with minor mode [29, 82]. Some studies suggested the existence of an interaction between melodic contour and rhythm, which can influence the perception of happiness and sadness [55]. This factor is highly context dependent, and some researchers found it to have little importance in emotional expression.

Harmony

Harmony is usually a question of chord consonance and dissonance. Simple and consonant harmony is perceived as dreamy [32], majestic [97], relaxed [55], happy and dignified [32, 97]. On the contrary complex and dissonant harmony is associated with anger [55], tension [81, 49, 64, 55], excitement and sadness [32, 97].

Depending on the implied harmonic intervals, chords can be classified according to their mode, major and minor chords. Similarly, major and minor chords are usually perceived as happy and sad, respectively. Cultural considerations affecting melodic intervals also apply in this context.

Tonality

The study of tonality and its dynamics is rather complex, in part because the concept of tonality may refer to different things. However, chromatic tonality is said to have been used in sad and angry melodies [90]. At the same time tonal melodies tend to be perceived as happy while atonality is usually associated with anger [90].

Rhythm

The term rhythm is used to refer all musical characteristics related with note's duration and their temporal organization. Rhythm is also a very complex factor usually studied from a general point of view. Research often analyses rhythm according to its regularity and flow. Regular/smooth rhythm is perceived as happy, majestic and peaceful [97] while irregular/rough rhythm expresses amusement [97] and uneasiness [31]. Happiness is associated with varied rhythmic [90] constructs. Flowing rhythm is perceived as happy and serene [32] while firm rhythm is perceived as sad [32, 98], dignifying and vigorous [32]. Gabrielsson and Lindström [28] point that the heterogeneous terminology among different authors, makes result comparisons difficult.

Meter

The property of meter is tightly related with rhythm, and like this one it is poorly explored and generically studied. In a study with children - 6 to 12 years old - Kratus [48], studied the perception of emotion in such young ages, in terms of rhythm- including tempo, activity and meter- modality, articulation, dynamics, interval size and use of imitation. Kratus classified each piece in terms of meter, according to the number of beats in each measure; duple meter if the number of beats is

divisible by two, or triple meter if divisible by three. The author applied multiple regression analysis in two data sets, one using valence and the other arousal as criterion variables. The results suggest that rhythmic activity and articulation are the two variables that best explain variations along the valence dimension, while rhythmic activity and meter explain variations in arousal, with duple meter being perceived as less exciting than triple meter.

Timbre

Different instruments may also influence emotional perception. A violin sounds very differently than a saxophone or piano, and it seems that those differences make some instruments more adequate to the expression of certain emotions. Research focuses on tone harmonics to explain those differences. Many harmonics suggest anger, disgust, fear and sometimes activity or surprise [82], while few harmonics seem to be associated with pleasantness, boredom, happiness and paradoxically sadness [82]. The absence of harmonics are usually perceived as tender and sad [38]. Although difficult to explain and study, the ability of certain instruments to better express certain emotions is rather intuitive and known by composers, performers and listeners. A fine example is the main theme of the 1993 film by Steven Spielberg "Schindler's list", written by John Williams and performed by Itzhak Perlman, featuring a crying solo violin able to surround the listener in the deepest sorrow.

Articulation

Articulation is an important property of music performance and extensively trained in classical music. Staccato is usually associated with energy, activity [98], fear and anger [38] while legato is associated with sadness, tenderness [38], solemnity [73] and softness [98].

Amplitude Envelope

In some instruments, notes can be played with different attack and decay times. A sharp envelope - rapid attack and decay - associates with anger [38], happiness, surprise and activity [82]. Soft or round envelope is associated with tenderness [38], fear [82, 38], sadness and disgust [82]. This is highly associated with music dynamics and likewise it is largely a performance property that also depends on the instrument.

Musical form

This factor is among the least studied; however, simple forms along with a moderate dynamism seems to be related with positive emotions [35], joy, peace [1] and relaxation [64]. Complex forms are perceived as sad [1] and tense [64, 49], melancholic and depressive if associated with a low dynamism, expressing anxiety and aggressiveness with high dynamism [35]. The usage of repetitions, rests, condensation and sequential development can increase tension levels [64], and tension peaks followed by rapid decrease, can be perceived [49] at the end of large sections. Disruptions of the global form are found to have little effect or no effect at all on emotional expression [30, 91, 46, 47, 42].

2.3.2 Final considerations

The above matter is gathered from a considerable amount of different studies, some of them dating from the first half of the twentieth century. As a consequence of the lack of definitions, terminological inconsistencies are not uncommon among those studies, making data analysis and comparison a much harder task. It is important to come up with a broad consensus on concepts and definitions, so researchers can talk and report its results in the same language.

Some structural factors need more attention and careful study including, melody and rhythm. The complexity and difficulty of these topics is probably the main reason behind such lack of attention. However, an effort to decompose and understand those structural factors is needed, so they can be appropriately studied and its role incorporated in the main knowledge base. Musicians may be of great help on such a task.

Music and its character emerges from a complex set of factors and interactions between them. It is unwise to think that a single structural factor may determine emotional perception. Emotional perception results from the interactions between all the above mentioned factors, and others omitted. Interaction relationships are virtually endless. Even if the interaction between two factors is somehow understood, the introduction of a third factor may render those principles inaccurate. Gabrielsson and Lindström [28] suggest that apparent data contradictions between studies, can be resolved by understanding the interactions between different factors. However, factor interactions are only part of the story. While it is true that structural factors are not static but rather dynamic, it is also true that listeners do not perceive them in a static way but dynamically as an interacting whole. Therefore, it is not enough to study each individual factor in specific contexts, and their dynamics inside a musical context should be considered to fully understand their individual role.

Not surprisingly, the ratio between questions and answers remains way above one; however, cooperation between researchers and a renewed interest for the area is likely to be the better opportunity to establish a solid knowledge base that can be used by a variety of disciplines to better understand music, emotion and their dynamics.

2.4 Rhythm Revisited

The word "rhythm" comes from the Greek, "ῥυθμός", meaning "flow", or "style" in modern greek. Rhythm can be defined as the variation in length and accentuation of a series of sounds. In a musical context rhythm refers to the duration of notes and rests, and the way they are temporally organised. Every musical note has a duration in time, ranging from milliseconds to seconds. From the time cadence of notes emerges rhythm, transformed in meter when subject to a rule, a measure usually denoted by a time signature. These three concepts can then be order in crescent abstraction level: duration, rhythm and meter.

Without different durations musical notes would only be distinguishable by their pitch, therefore, rhythm is an essential part of music. In fact, some authors and musicians consider rhythm as the very

essence of music; however, historically this role was not always recognised by Musicology and music theory. For cultural reasons the study of pitch has prevailed over durational questions. Rhythm and meter were attributed a secondary role in music structure, constrained to the impositions of melody and harmony. This bias seems to occur in music education too, i.e., new musicians are taught to consider things that way. Being the lucky owner of a "perfect pitch" hearing is far more valuated than being the owner of a carefully calibrated ear able to distinguish the slightest difference in duration, and naturally organise a whole music line.

The composition algorithm proposed in this work, AMADEUS, is for the most part based on rhythm. Although section 2.3.1 already provided some insight over its influence on emotional expression, that insight can not be taken as complete, since rhythm properties were studied superficially - surely this is related with its complexity. Thus, it is clear the need for a deeper and solid perspective on rhythm that somehow considers its perception qualities and upon which AMADEUS can be based.

Such perspective is naturally contributed by a percussionist; in his PhD. thesis Eduardo Lopes [56] develops an extensive theory of rhythm and meter providing adequate empirical support. The concepts of *Pulse Salience* and *kinesis* introduced by Lopes, provide the necessary theoretic foundations upon which the rhythm composition module of AMADEUS can be based, but this proposes two challenges, to provide the necessary formalisations and quantification mechanisms for pulse salience, and use an essentially analytical theory as a composition tool.

2.4.1 Pulse salience

Neuropsychological studies by Isabelle Peretz and José Morais [71] provide strong evidences that the perception of melody and rhythm occurs independently in a first stage, the two properties are only assembled later in the perceptive process. By studying brain injured individuals, they were able to observe that rhythm and melody perception occurs in different brain regions, implying the existence of different perceptive mechanisms. Supported by these strong neuropsychological evidences Lopes suggests;

> "If listeners do in fact segregate the durational parameters of music from the pitch parameters at the beginning of the cognitive process, it is plausible that they also attribute to each a different perceptual weight (what I shall refer to as salience)."

According to the author, pulse salience is a sum of three factors: metric placement, agogic accentuation and rhythm cell accentuation. The metric placement of a pulse relates directly with the meter "stratum". It is agreed that some metrical points are more stable than others - this is taught in music theory - e.g., in a quaternary meter the first beat is the most stable, followed by the third, second and finally the fourth. Placing a pulse in a stable position will have an accentuation effect rising its salience. The more stable a metric position the more salient a pulse on that position will be. In practical terms, if a pulse is placed in a stronger metrical point then it will be more salient than other placed in a weaker point - considering metric placement only. Agogic accentuation is directly related

with the rhythm "stratum". The simplest of three components is proportional to pulse duration, i.e., the longer the pulse the more salient it is. Rhythm cell accentuation occurs in the rhythm "stratum" also. The preceding rhythmic cells can influence the overall salience of a pulse. Specifically, depending on those cells pulse salience can be increased through accentuation. Lopes [56] defines a rhythm cell as a group of notes made in accordance with a beat framework, that tend to have a beat duration. For example, in the same quaternary context two eighth notes- forming a rhythm cell – preceding a quarter note, have an accentuation effect over the last one. This accentuation occurs due to the fact that, duration of each eight note is half of the quarter's duration.

2.4.2 Kinesis

Surely the reader already experienced some kind of motion when listening to music. For example, the third movement - Rondo Alla Turca also known as Turkish Rondo - of Mozart's Sonata N.11 in A major, is perceived as a high kinetic piece with an impetuous character. On the contrary, Vivaldi's Adagios for violin are perceived as calm, serene and gentle with low kinesis. The Brazilian Samba and its African roots are another example of extreme kinetic genres, as opposed to other South American musics such as, the passionate Argentine Tango.

Lopes, references the work of Neil Todd [92] to isolate music's source of motion. Todd identifies two different types of movement, gestural and locomotive. A gesture is defined as a single continuous movement, usually associated with expressive music and emerging from musical phrases - although he also refers that even a continuous sound can draw a simple gesture. Locomotive motion is associated with tempo, and assumes a prominent role in highly metrical music. Gestural motion can be experienced in arrhythmic contexts such as, Gregorian Chant, thus it does not depends exclusively on rhythm. Locomotive motion is the exclusive product of the action of rhythm and meter. Thus, gestural motion can be as free as immeasurable in arrhythmic contexts, but quantifiable in rhythmic music.

Lopes suggests that meter is music's main source of kinesis, with stability and kinetic potential correlating inversely. Placing pulses in weak – unstable - and highly kinetic metrical points - contradicting natural metric organization - releases kinetic potential creating a perceptual tension, responsible for the perceived motion. Tension levels can be resolved by returning to a natural placement of pulses, i.e., placing a pulse on the next strong beat where resolution is supposed to occur.

2.5 Computer Music

In 1957 a revolutionary article received enthusiastically by some and with doubtfulness by others, was published in Scientific American. The article by Lejaren Hiller and Leonard Isaacson presented the *"Illiac Suite"*, a piece for a string quartet sometimes called of *"String Quartet No. 4"*. The *"Illiac Suite"* was integrally composed by the "ILLIAC I", a computer built in 1952 by the University of Illinois, USA and the first to be entirely built and owned by an educational institution. The *"Illiac*

Suite" is considered the first musical composition created by a computer. Hiller commented;

> "I observed that if we could program a computer to simulate a 'walk' through, say, ordinary space, we could also simulate a 'walk' through a grid defined to represent musical elements such as pitch, rhythmic durations, and timbre choices ..."

The *"Illiac Suite"* results from Hiller's and Issacson's pioneering studies in computer music. The two scientists were also musicians wishing to explore the artistic potential of the computer. The main objective was to test at what degree rule-based algorithms could be implemented to compose style or genre dependent music. The *"Illiac Suite"* was for the most part composed using traditional composition rules but its fourth movement was composed using mathematical principles, namely Markov chains. As a curiosity, by that time the computer was not able to output the sound itself, the ILLIAC I was only able to output computation results in five holed paper tape or using a CRT display. For that reason, the results were manually transcribed to common music notation and eventually performed by real musicians.

Taking natural advantage from advancements in computer systems, computer music suffered a great evolution since the *"Illiac Suite"*. Eventually, research reached a maturity level in which different methodological approaches co-exist. Although, it can be argued that the area remains in its "teen" years. While this section provides some insight over the most popular approaches, using practical implementations of composition systems, it is not in any manner exhaustive, a broader overview and deeper discussion is provided by Roads [76]. Nevertheless, this section should frame this work in the area of computer music. Modern research on music composition is usually classified (e.g. [37]) according to its implementation techniques: rule-based, stochastic, grammar-based, genetic algorithms and neural networks. Recently, genetic algorithms seem to have taken the lead, and most of the following discussion is centred in such approach. In addition, two important projects using neural networks are mentioned separately in this section.

2.5.1 Rule-based composition

Rule-based systems try to capture composition techniques, taught and learnt by real composers, in a set of computation rules to which the final composition must conform. This type of approach is probably closer to how real composers do their work. In spite of their artistic subjectivity, composers systematically apply a set of basic rules ensuring good and ear pleasant results. The specific rule formalisation is not relevant, the common denominator between this approaches is the use of composition rules usually derived from the music composition discipline. Some rule-based techniques consider strict rule adherence but others contemplate a determined degree of freedom inside which the result is considered valid, e.g., Schottstaedt [84] used a penalty system to attribute a different level of relevance to each rule.

The Soundtrack of Your Mind

Eladhari et al. [22] propose an experimental system aiming to increase believability and immersion in computer games. The Mind Music system was developed in the context of a computer game - Garden of Earthly Delights - in which the player is exposed to three music sources: ambient sound, situation specific melody themes such as leitmotivs for objects and rhythmic variations to express the level of energy/excitement. The system implements a mind module based on the affective paradigm, that can be applied to the games entities or characters. The idea is to use affective information from those entities to dynamically adapt generated music.

The authors define a mood matrix consisting of two mood scales with a total of 25 cells, given that each scale has five hard segments. Depending on the activated emotions at a given time only one of these cells is active. Music factors, Musical factors, harmony and time signature, vary in each of these cells. The mood scales represent two mood circles, the inner mood representing character specific and private feelings and the outer mood reflecting the relations of the character with the game world and towards other characters. The inner mood is represented by the horizontal scale of the mood matrix controlling the parameter of harmony, while the outer mood in its turn is represented by the vertical scale of the mood matrix controlling time signature.

The resulting sounds depend on the different modulations – twenty five - that can occur inside the mood matrix. Each modulation is a MIDI sequence built with Microsoft's DirectMusic Producer [61]. For each emotion there is a characteristic melody that cannot interfere with the harmonic structure and time signature. Harmony and time signature need to be independent because they will represent the two different moods; however, the composer maintains some degree of flexibility when composing these small leitmotivs, e.g, it can change the instrument, note density and sound effects without interfering with harmony. One important observation is that these melodies should work with all possible modulations of the mood matrix, relaxing the need to define multiple melodies for each possible harmony.

Ka-Hing et al.

Ka-Hing et al. proposes a music arrangement system [41]; the authors aim to develop a system able to provide computer games and interactive applications with meaningful orchestrations depending on the affective requirements of such applications and games. The idea is to provide a single interface that can be used to input music materials, rules and parameters by the composer and game states, their transitions and associated emotions by the game-designer. The authors separate musical data, arrangement rules and the arrangement process itself in a layered architecture. Rules define music qualities, such as rhythmic patterns and tempo, that should be used in a particular music intended to express a given emotion. Different arrangement styles can be defined using different sets of arrangement operators at the arrangement process level, which may imply the definition of specific rules associated with a certain style.

Scherer and Zentner production rules

In a slightly broader concept of "rule" Scherer and Zentner [83] present a set of rules to predict listener's emotional response in the presence of certain musical characteristics. Inclusively, the authors suggest the use of these production rules in a computational model, although they warn that current knowledge is not precise enough, to precisely model the parameters and input/output variables. Section 2.3 discussed the difference between the perception of emotion and production of emotion in listeners, without prejudice of that section Scherer's and Zentner's work is referred here as an example, of a rule-based approach, to the production of emotion in listeners.

2.5.2 Stochastic composition

Probabilistic approaches were very common among early research, e.g., the usage of Markov Chains in the fourth movement of the *Illiac Suite*. Probabilistic models based on Markov chains are indeed common. Among others, Farbood and Schoner [23] used Markov chains to compose music known to be aesthetically good and pleasing.

Microsoft DirectMusic

The IT giant, Microsoft Corporation, develops a system for music composition and manipulation intended for usage in computer games and multimedia applications. DirectMusic [61] is distributed as a module of DirectX, the famous SDK from the Redmound's company. The framework includes a tool called DirectMusic Producer, designed to interact with the DirectMusic module. DirectMusic Producer uses a probabilistic approach to describe chord progressions, and provides interactive functionalities - using the so called IMA, Interactive Music Architecture - that can be used to implement interactivity between the music and the user. Combining this functionality with some musical base - user provided - it is possible to compose music in real-time, adapted to whatever the programmer wants including character emotions, if so desired. DirectMusic Producer is a complex tool not suitable for the casual musician or programmer having a steep learning curve. DirectX as well as all its modules and components are proprietary software of Microsoft Corporation, turning its usage for research purposes less attractive.

The Affective Remixer

The Affective Remixer [5] is a real-time arrangement system observing the listener's emotional state and reacting accordingly. The system infers the current emotional state through a probabilistic state transition model, implemented with Markov chains. The idea is to record listener's reactions to selected music segments, and from there infer about the affective potential of each musical parameter. This information is then used to re-arrange music segments according to affective cues provided by the listener, and recorded in real-time. The authors define a stimulus matrix organising each music segment by parameter. A specific arrangement is defined as a path trough the grid that incrementally

appends segments to the arrangement. To represent the way a certain musical parameter changes the listener's affective state, the author's used two bi-dimensional models, one for the affective states and other for the audio segments, variations are represented through vectors. Affective states are classified according to their valence and arousal, horizontal and vertical dimensions respectively, while audio segments are organised according to their rhythmic complexity and layer thickness, horizontal and vertical dimensions respectively.

The affective state transition model is in fact composed by two sub-modules, affect-detection and action-chooser, both deriving from the data collected in the initial recording of listener's reactions, and both based in Markov chains. The affect-detection model is responsible to analyse listener's input - foot-taping and GSR (Galvanic Skin Response) - and infer its current affective state. Given the current affective state and a target state, the action-chooser module is responsible to trace a path inside the music matrix that most likely will lead to the desired affective state.

2.5.3 Grammar-based composition

Context-free grammars have also been explored for music composition purposes. Using a grammar for composition usually requires basic musical construction elements such as, rhythms, to be described in terms of grammar symbols. Together with grammar "rules", it is possible to manipulate symbol sequences to respect those rules, representing admissible element sequences.

McCormack

McCormack [59] uses string rewriting grammars for music composition. His work is based in a special type of formal grammar known as L-System - Lindenmayer systems - which are extensively applied across many computation fields, such as computer graphics, biological systems and morphogenesis. L-Systems are used to represent the repetition and self similarity of musical phrases. McCormack extends the formalisation to simulate the probabilistic characteristics of Markov chains, effectively introducing a probabilistic measure in the genesis of admissible phrases.

2.5.4 Genetic algorithms

Genetic algorithms became very popular among the research community. They have been applied to harmonisation [99] and jazz [24], for instance. Some of the most relevant computer music systems have used genetic algorithms for composition. Somehow reflecting this recent popularity, this section is larger than the ones dedicated to the other three approaches. However, before starting to analyse each study and since all of them share the same concepts, it is perhaps useful to say a word about Genetic Algorithms and their origin. Sometimes, variations and extensions exist, in such cases and when relevant those variations are presented in comparison with these base definitions.

Genetic Algorithms - shortened to GA in the text from hereafter - are global search heuristics used to find exact or approximate solutions in optimization and search problems, being part of a broader

family of algorithms referred to as Evolutionary Algorithms. GA are inspired on the Darwininan theory of evolution and were extensively developed by Holland [34]. In fact, for people familiarised with biology in particular with genetics, the concepts of GA would sound very close and familiar. In GA a solution is a collection of "chromosomes", each made of a variable number of "allele", the atomic unit in GA. Genetic changes in the chromosomes - evolution – occur by the action of two basic operators, "crossover" and "mutation" – mimicking real natural processes. The selection mechanism, responsible to select the fittest and extinguish the unadjusted exemplars - solutions - is captured in the concept of "fitness function" - the natural selection process of Darwin. The fitter solutions as assessed by the "fitness function" are given a chance to reproduce while others are gradually extinguished. GA are applied iteratively over populations of candidate solutions until a stopping criterion is met. The iteration cycle is a 5 step process: the initial population is usually randomly generated, the fitness function is applied over the initial population, based on this evaluation a set of parent solutions are selected, the operators of mutation and crossing are applied on the selected set to produce new chromosomes, the fitness function is applied over the candidates of this n-th generation and the algorithm returns to step 3 until the stop criterion is met. After this introduction, what follows is a non-exhaustive list of some relevant studies using a genetic approach to the problem of music composition.

Numao et al. 2002

Numao et al. [65] develop a composition and arrangement system based on user's affective response.

In a first stage user's feelings towards selected music pieces are recorded and analysed, from this data emerges common musical structures, eliciting a specific emotion. These structures are extracted by combining Inductive Logic Programming - a machine learning technique to find PROLOG rules - and elements from tonal theory. The set of logic inducted conditions represent a personal model that can automatically evaluate a chord progression, i.e., such model can be used as the algorithm fitness function. The fittest progressions as selected by the GA, provide the arrangement basis, while composition is done through a melody generator using those same progressions.

Unehara et al. [93] propose a similar approach; the difference resides in the way personal and subjective data is collected, Unehara uses direct continuous user interaction while Numao, uses a rather refined and complex methodology, reducing user interaction to the initial training of the fitness function.

Very recently, Legaspi et al. [52] presented in the IUI'07 conference, a specification for a music composition system based in the listener's affective perceptions. The approach and architecture derives from Numao's work - Numao is also an author of the referred article - using the same inductive logic programming paradigm to describe the relations between structural features and emotions. The inferred model is then used as a fitness function in the GA backing the music composition process. Due to the similar philosophy and tight relation between the two works, both were referred in this same section.

CBS: A Concept-Based Sequencer for Soundtrack Composition 2003

Jewell et al. presented in the WEDELMUSIC'03 conference [37] a composition system referred by the authors as a Concept-based Sequencer. The main goal of the project is to combine semantic markup of temporal media (e.g. film, radio) with GA to compose music that can fit the scripted annotated content. The CBS concept model is supported by two basic concepts or node types, concepts representing any named entity in the system (e.g. item, attribute or mood) that can be hierarchically organised, and events representing a significant occurrence within the analysed media, evolving several participants and concepts delimited in time. A GA algorithm is responsible for the evolution of the generated musical sequences, subject to the fitness function assessment. Each musical note corresponds to an a allele in a chromosomal string, a similar approach as that of Wiggins and Papadopoulos [99]. The fitness function in the CBS system is based and highly dependent on modifiers introduced in the composer mappings. The concept of composer mappings can be used to impose some "musical modifiers" over the conceptual representation, i.e., the conceptual interpretation may change between composers and somehow the system is able to reflect these differences in the composition process through composer mappings. The authors claim that this extension to GA, allows for a flexible interpretative composition.

Unehara et al. 2003

Unehara and Onisawa [93] propose an Interactive GA approach. In Interactive GA the evaluation process, and therefore candidate selection, is based on some type of input, usually user input, instead of fitness functions. The authors justify the option with the claim that fitness functions are very hard to conceive in certain areas, such as arts and therefore music. This is indeed true, artistic subjectivity is hard to capture. The idea is to bring up that subjectivity directly from the users at no cost for the system. In this work each chromosome is a four bar musical sequence, a whole piece is a set of four chromosomes individually, i.e., sixteen bar sequences. The user feeds the selection mechanism by evaluating each chromosome. One of the goals is to help non-composers to compose their own music reflecting their subjective evaluation.

Birchfield 2003

David Birchfield proposes a generative model for music composition. Birchfield's work is supported by emotional theories and meaning in music, as well as cognition and perception research. The author discusses two common approaches to generative models. The first is based on the systematic analysis of musical styles providing a generative basis for the composition of new musics, that share the same stylistic features [10, 11]. Birchfield points out that, while this can capture formal and grammatical features of a given style, it is impossible to explore new grammars of yet unknown music. The second approach is to use existing generative models, such as cellular automata and L-Systems and experiment with different data [63]. For Birchfield this can lead to interesting results, but since composed music

is bound to the patterns of the input data, results may not be "truly compelling".

Alternatively the author aims to build a generative system that can produce "compelling results" for one side and be stylistically independent and unique in the other. His approach is highly influenced on the works of Meyer [60] about emotion and meaning in music, and Lerdahl [53] about composition style versus listening experience. Birchfield's generative model is based in a co-evolutionary GA, with a population made of hierarchical musical components, navigable and discernible by the listener. The main idea is that independent trajectories for each musical component produce large-scale forms. The selection of such components is based on music cognition and perception, to ensure intelligible musical results. The author sustains his option with the flexibility allowed by GA, suited for musical manipulation and enhancement. The power of combinatorics is able to provide unpredictable and spontaneous outcomes which is in accordance to the author's objectives, produce novel results from which new forms and musical ideas might emerge.

Kim et al. 2004

Kim and André [43] propose a music composition system for the controlled manipulation of the user's emotional state. The user is not required to continuously provide feedback to the selection mechanism. Such as Unehara et al. [93] and Numao et al. [65] the authors use user-specific data to evaluate candidates - affective data in this case. Kim's and André's system uses two stages to come up with a fitness function. Initial training is done through explicit user judgements, in same way Numao's does, and once the system knows the user and its emotional responses, the fitness function is drawn on physiological feedback collected through sensors. Inferences on the user emotional state are drawn by a rule-based system combining sensor input with empirical data, and emotion is represented in a bi-dimensional space of valence and arousal.

2.5.5 Neural networks

ADA 2002

In 2002 one of the attractions present in the Swiss National Exhibition Expo'02, caused a big impact on visitors, serving as a showcase for the state of art in multimedia interactive systems. The attraction causing this impact was, ADA: Intelligent Space [96].

ADA is an artificial being able to communicate with humans through emotions. ADA's sensor system, included fourteen video cameras, six microphones and almost four hundred pressure sensors in floor tiles. Output was done through lights, real-time computer graphics projected in a three hundred and sixty degree canvas, and real-time music composition and performance. ADA is a perfect example of integration and hybrid software architectures, combining standard procedural programming was combined with agent-based methodologies and neuronal networks.

The goal was to impel visitors to adjust their behaviour to ADA, using both the visual and auditory channels. The auditive channel was attributed to the Roboser system [95, 36], a real-time

music composition and performance system, able to accept input from a variety of sources and hence, suitable for integration in a variety of systems. In ADA, Roboser was fed with simple events detected inside ADA's space, e.g., footsteps, whistles and hand claps, used to express parameters, such as joy, arousal and surprise. Music performance and composition was synthesized on two separate layers, mood and emotion, reflecting ADA's control model. Roboser is able to compose music as a MIDI data stream with up to twelve tracks. In ADA, two of those tracks were used for mood composition, expressing arousal and valence, while eight tracks were used in the emotional layer to express four basic emotions: joy, surprise, sadness and anger, for a total of 10 tracks. At each moment, one emotion's tracks were set to maximum volume, the others remained muted, and changes between emotions were made gradually.

The results achieved in ADA were very encouraging, the music generated by ADA was used on a documentary film by the Swiss Federal Institute of Technology whose aim was to introduce the general public to ADA's concept. The documentary *"Brainworkers"*, won the award of Best Swiss Commissioned Film of the year 2002, including a win in the sound design category. The authors claim that ADA passed the musical Turing test, because the jury was unaware that the Brainworker's soundtrack were for the most part composed by ADA.

VR-Roboser 2007

Sylvian Le Groux et al. [51] apply the Roboser [95, 36] system to compose real-time background music in virtual environments. VR-Roboser was tested on a dynamic avatar able to explore a virtual three-dimensional environment, and whose avatar behaviour influences the generated musical events.

The system is decomposed in a three stage process: sensing, processing and responding, each implemented in independent modules. The sensing module consists of the virtual environment and the avatar exploring it autonomously, the processing model is a neural network simulator that analyses sensed data, while the responding model composes the music whose structure is influenced by the avatar's behaviour. This third module is backed by the Roboser composition system. The input of Roboser can contemplate raw sensory data (e.g. video, audio, events), biased neural oscillators, models of circadian rhythms or behaviour control. In VR-Roboser it receives input from the IQR controller implemented in the sensing module. IQR is a simulation environment for large-scale neural models providing tools to build and run such models. The IQR module handles real-time input from the avatar's sensors. Music is composed in a multi-track MIDI stream and the parameters that can be independently manipulated in each track include; instrument, velocity, volume, pitch bend, tempo and articulation. Roboser uses some predefined templates such as, music fragments, rhythm lines and Dynamics- note onset or loudness- that can be selected for each track, and stored in a "Style File" for later real-time interaction.

2.6 Film Scoring

Since the born of the seventh art, music has been accompanying film. The word film, refers to the narrative dramas commonly found in cinemas, television or video. Common to the experience of film and music there is emotion. It is therefore pertinent to ask in which way, music, on its intimate relation with film, contributes to the emotional aspect of narratives. The topic received little attention until the last decade of the twentieth century by the hands of Prendergast [72], Cohen [6] and Marks [58] the topic began to appear in the Psychology area. Film scoring is a practical application of story sonification with expressive and communicative intents. Considering the present work's motivation, namely I-Shadows as an Interactive-Drama application, film scoring is worth to look at. The perspective on Film Music provided in this section, is firstly psychological and based on Cohen's contribution [8].

Cohen suggests that the study of emotion in film provides, in general, greater terminology clarity. She exemplifies with the concepts of emotion and mood, usually distinguished by the idea that emotions need an object towards which they are projected, and mood does not. In fact under a film context music attaches itself, to the objects focus of attention, or implied narrative topic. Such attachment is characteristic of multimedia environments and it is usually much harder to identify such objects in a pure musical context. Cohen argues that film psychological studies can be useful on questions about music and emotion, while emotion and autonomous music knowledge can help the study of the processes occurring in film context.

2.6.1 A brief historic perspective

In the early times of film the usage of music had a very pragmatic primary function, to mask the projection machines noise. Of course music was not used exclusively to mask projector's noise, it was also explored to improve narrative intelligibility compensating the absence of dialogues. Eventually, the problem of noisy projectors was a brief reality and projectors become more silent, but music remained, and a whole music film industry developed. Back on those days music was live performed, record techniques would only appear later in the twentieth century. This caused a huge increase in the procurement of pianos to equip cinemas, whose architecture contemplated special spaces dedicated for soloist musicians or small orchestras, and music anthologies featuring a music repertoire for the expression of various emotional settings were edited.

The new art form was a revolution in entertainment and had a strong social impact. The first scientist to address such changes was Hugo Münsterberg. Münsterberg was completely amazed by the emergence of this new aesthetic form, and the possibility to study its genesis. In that time, and although Münsterberg own experience with film was very recent, he draw some conclusions about the new fascinating world. About music he noted that was a tension relief mechanism, helped to maintain the interest on the action, provided some comfort and reinforced emotion, besides having a strong contribution for the aesthetic experience.

The year of 1927 saw the appearance of "The Jazz Singer", the first film to incorporate voice and sound effects. With the introduction of voice in films, the first impulse was to demise music, since it was no longer necessary to support the narrative in which dialogues could now be incorporated. However, interesting enough, was the immediate loss experienced by film watchers. By that time it became clear that music was an essential part of film. Münsterberg claimed, that as an art form film is much more related with music than to painting, with which shares superficial resemblances only. Authors such as Palmer [68] and Rosar [77] suggested that music is able to add a third dimension to the bi-dimensional projection screen. This view is, of course shared by music composers [12].

2.6.2 Functions of film music

Different authors point different functions for film music; however, there is some level of agreement for the most part. This section presents three complementary perspectives on film music, one by a psychologist, a film editor and a music composer, in this order.

Annabel J. Cohen

Cohen's [7] influential work identifies eight functions of film music. The first function is to mask extraneous noises that can interfere in the watcher's attention. This was the most important function in the early days of cinema. Secondly, music provides continuity between different camera shots informing the watcher that action has not changed in space or present characters. Third, music has the ability to direct attention to important features of the plot; this is achieved through structural or associative congruence. When dissociated of objects music is able to induce mood, being this the fourth function identified by Cohen. Music can effectively communicate meaning, and thus it can be used to further the narrative and provide a resolution tool for ambiguous situations. Music is also able to symbolise, past or future events as well as specific objects or characters through the usage of the leitmotiv technique supported by an associative memory mechanism where such associations are stored. Each event, object or character has an associated leitmotiv, that is present whenever the associated element is present. The sense of reality can also be increased by music deepening the absorption of the watcher, perhaps increasing arousal. Finally but not last, music has an important contribution to the overall aesthetic effect of film.

Oppenheim, Y.

As a film editor Yair Oppenheim [66] purposes a more practical list. Providing necessary film music; there are different types of necessary film music, the most important being referred as "source music". Source music exists inside the film, and it can be heard by the characters; for example, if there is a character playing a piano its sound should be heard, or if there is a radio on the plot, there should be some music coming out of it. The simplest form to provide source music is to embed songs into the scenes at the appropriate time, while the most complex form is to compose orchestral music

specifically for the film. The most complex case is when the entire film is based in a single score- the film *Close Encounters of the Third Kind* by John Williams is a rare example. Other important types of necessary music are the opening credits and character themes. Opening credits music should tell the viewer what to expect from the story and follow the action on the screen, e.g., the *Star Wars* opening credits music reflecting what the watcher sees in the screen, a wide view of space. The themes associated with characters, are also an example of how music should match the action, and especially important when there is a strong identification of each character with an archetype, such as the hero or the villain.

Identification of ethnicity, location and period; this is a very important function for the action's historic and social contextualisation. The simplest method is to use a pre-existing composition, reflecting the action's location and period. When specific music is composed it must retain the same aesthetics and characteristics representing the location and period. Usually, this requires specialised study and training by the composer. *Braveheart* from 1995 is a fine example where the composer James Horner, contextualised the action using bag pipes for the instrumentation base.

Paralleling action; this is also known as underscoring, and refers to the usage of music as an accompaniment for the action. Rather than summarising the overall mood of a scene, it strictly follows the action. Music underscoring is a "weak" function, and so, if used incorrectly it can produce weak results. The film *The Rock* of 1996 is such an example where it is impossible to musically distinguish between scenes. Some composers follow a strict cue by cue matching, which requires quick and fluctuating rhythms and atonal music. A simple type of underscoring is the inclusion of the character's theme in every scene he appears on, saying who is on the screen at each moment. Usually the character's theme has a number of variations, so it can adapt to different scenes. However, this only works if there is a single character on scene or the film action is centred on it.

Commenting; when music adds something to the scene, additional information or even some considerations and insight. This type of function can be achieved by more than one mean, overtures are among the most important. An overture must include all the composition's basic themes, and it should provide a summary of the plot, e.g., the overture of *Superman* from 1978 includes a march, presenting superman as the hero, and a love theme, the hint of the existence of a "princess" to rescue. Other important type of commenting is location description, e.g., the film *Legends of the Fall* from 1994 has an opening scene featuring a panning shot of a landscape. James Horner, used string instruments to produce sweet melodies reflecting the beauty of the landscape, and helping the viewers to enjoy such beauty. Sometimes, characters evolve during the action and that can be represented in the evolution of its theme, e.g., recalling again the film *Star Wars*, Luke Skywalker is initially depicted with a vigorous and energetic orchestration reflecting his youth, at the end, the same theme uses trumpets and acquires the character of a march expressing the maturity and strength of the young hero. Luke's theme evolved helping the viewers to perceive the character's evolution.

Produce emotional responses on viewers; this is one important function of the film score and it is usually achieved trough the means of different orchestrations, i.e., using different instruments, e.g.,

strings emphasise romance, serenity or tragedy as in the film *Schindeler's List* from 1993 by Steven Spielberg, brass instruments emphasise power and percussion is used in crescent cadences to create suspense. Some critics believe that this is the only true film score function, because all other functions try to produce emotional responses on the viewers.

Aaron Copland

In a more generic way the composer Aaron Copland identifies four interactive relations between music and film. Music can be used to create convincing atmospheres of space and time, underline psychological refinements of the story, fill the background and build a sense of continuity, underpinning the theatrical build-up of a scene and rounding it off with a sense of finality.

Chapter 3

Proposed Architecture

"The jar shapes the emptiness, such as music does with sound."

Georges Braque

The development of composition algorithms, usually requires the implementation of additional computer programs that can support the execution of the compositional procedures, e.g., audio output modules. The design and implementation of such support logic, often shifts the researcher's efforts into fields other than computer music, accounting for a significant part of the available research time. Research efficiency can be improved if the researcher is able to take advantage of a programming interface provided by a development framework and concentrate in the development of the composition algorithm.

This chapter proposes a software architecture for an emotion-based composition algorithm development framework and run-time environment. While, this architecture was designed to support the development of AMADEUS, it aims to go a step further and address the problem mentioned above. This architecture together with its implementation - see chapter 4 – aims to be a useful research tool to assist the development and implementation of emotion-based composition systems that can be easily integrated in a variety of contexts. Researchers are expected to use these tools as a programming environment upon which they can prototype and test their theories, thus, this proposal is rather flexible and extensible, postponing context dependent decisions. The proposal consists of four parts: the first part discusses the layered structure of the algorithms run-time environment, the second part defines the application programming interface to interact with the that same environment and develop the composition algorithms, the third part provides brief guidelines for the system's integration in affective applications, finally, the fourth part discusses optional architectural components providing additional programming support.

However, before discussing these elements, this chapter presents three fundamental high-level requirements and the design guidelines, that strongly influenced the architecture's conception. In the

remaining of this chapter the term, "affective application" is used to refer the virtual environment in which the composition system is integrated, the composition system implementing this architecture is simply referred as "system", and the word "programmer" refers to the researcher using an architecture compliant composition system in his research activities.

3.1 Requirements and Design Guidelines

Generality and extensibility

Chapter 2 discussed a number of different approaches to computer generated music, ranging from rule-based composition to genetic algorithms. These approaches require the usage of different formalisations, models and impose different requirements. In its turn, affective systems may also use different conceptualisations, e.g., emotion models. Thus, a general architecture should be flexible enough to accommodate each approach's specificities, but an increased level of generality usually means poor optimisation. Any serious attempt to provide a general tool, must include extension mechanisms that can provide missing and custom functionality. This work's strategy is to provide basic extensible functionality and models, that can be optimised to comply with domain specific requirements, which could not have been considered before.

Real-time processing

Using music for interactive purposes usually imposes real-time constraints. Indeed, it is important that appropriate music reaches the user's ears at the correct moment, never before and definitely not after. Using untimely music may in fact decrease interaction quality, which is precisely the opposite of what it is supposed to do. In this context, real-time constraints are in fact soft real-time constraints, i.e., failing to compute and deliver a particular segment at a particular time does not imply system critical failure; however, it has a negative impact on interaction quality. This negative impact can be minimised if the system implements contingency mechanisms taking adequate actions on job failures.

For composition systems the most critical activity to which these real-time constraints apply is essentially the composition stage. While these constraints should be primarily addressed by the programmer, by implementing efficient composition algorithms, the underlying support logic must provide real-time performance auxiliary processes, that although not part of the composition algorithm itself, are used by this one.

Therefore, the implications of these constraints on the proposed architecture are essentially related with the composition and output stages. It is important to ensure that a particular element receives the relevant data at the right time. If the system fails in doing so, then it should ensure that the now irrelevant data is discarded, and room is made for the next block, avoiding the "traffic jam" effect. These considerations imply the continuous monitoring of the data stream to detect delays and immediately correct them, without affecting data flow considerably. This also implies transporting data along its path inside time aware elements, which essentially have a "life time" stamp representing

the time interval the transported data has to complete the processing chain.

To achieve real-time performances, the composition stage should be divided into rather independent tasks, requiring concurrent programming and data access synchronisation. Additionally, and since the control over the exact rate at which each processing task executes is limited, buffering mechanisms must provide temporary storage for data waiting to be processed by the next task in the processing chain. These buffering elements should provide non-blocking write, read and modification operations, i.e., if a certain task can not be performed, the respective thread should release all synchronisation resources and eventually try again later. This is a crucial aspect to avoid starvation and dead-locks. To reduce system complexity and improve performance, the synchronisation logic should be managed by the buffering elements themselves, i.e., synchronisation should be transparent to the calling task. If buffering capacity is exceeded, the most recent data should receive higher importance, because the oldest has a higher probability of exceeding its life time during the processing chain, and thus it should be discarded.

High system integrability

Considering the myriad of development environments and tools as well as techniques, it is important to provide a flexible integration mechanism. Although there is a set of common tools and programming environments, more exotic combinations may appear, being therefore important to consider various integration options. The challenge is to define a common model that can account for various integration options while ensuring a detail independent and coherent assembly. It is important to note that integration does not exclusively concerns the communication between the affective application and the system. Integration, usually requires concept translation, when conceptual differences between the application and the system exist. Therefore, it is not only a question of programming environments and languages, but sometimes a question of integrating different models and development philosophies. These integration questions should be treated independently from the normal interaction between systems, i.e., differences and translations should be transparent for both the affective application and the composition system.

3.2 Run-time Environment

Composition systems can be structured in terms of three different stages: sensing, processing and response [78, 94]. Some architectures reflect this decomposition explicitly, through the implementation of independent modules, one for each stage, e.g., VR-Roboser [51] mentioned in section 2.5.5. This approach offers some important advantages. The first is complexity management, the problems of sensing, processing and output are very different in nature, usually requiring different implementation models, designing the solution for each of these problems in a rather independent manner reduces the system's complexity. On the other hand as long as an interaction interface is previously agreed and kept stable, module development and testing can be done incrementally, or in parallel in the case

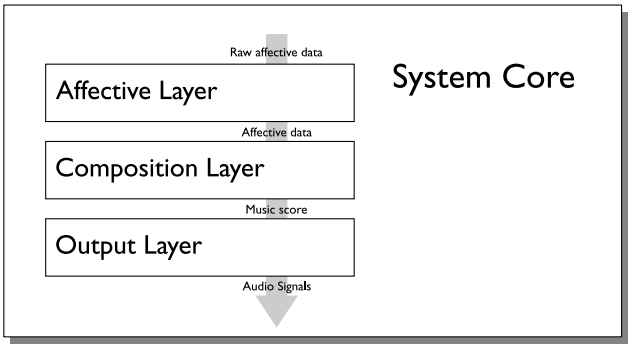


Figure 3.1: Layered structure of the run-time component

of an extended development team. Finally, future extensions inside a single module do not affect the other two. This proposed architecture follows the same paradigm and consists of three processing layers: affective, composition and output, corresponding to the sensing, processing and response stages respectively. Figure 3.1 shows the layered structure of the architecture's run-time component.

3.2.1 Affective layer

The affective layer - figure 3.2 - manages an intermediate state representation of the affective application – it can be seen as big affective buffer. Such intermediate representation is required due to the real-time requirements of the composition process, the affective data needs to be readily available. Maintaining an intermediate representation of the affective application's state introduces the additional requirement of ensuring coherence between representation and the original data, evolving in the affective application. To avoid exchanging additional synchronisation information, the system should not allow other image modifications than the ones resulting from the affective application's update requests. In this sense, the internal representation is a static snapshot, i.e., it is only updated upon request from the affective application. While there might be recurrent operations performed after certain events, the above considerations imply an explicit request from the affective application to perform these operations. However, performance would be improved if it was possible to delegate these recurrent operations over the composition system, i.e., these operations are implicit in the main request of the affective application.

The **Affective Environment** is the control component and the interface of the affective layer. The affective environment manages all the affective entities and their emotional states and it is responsible for delivering relevant affective data to the composition layer. The environment provides a picture of the overall environment's mood in a single value. The type, meaning and concrete quantification of such value is subject to specific implementations.

Control politics continuously monitor the affective representation and decide what data, and at what time, should be delivered to the composition layer. The system maintains a set of different control politics, but at a given moment only one control politic may be active. At least one control politic must be present to run the system.

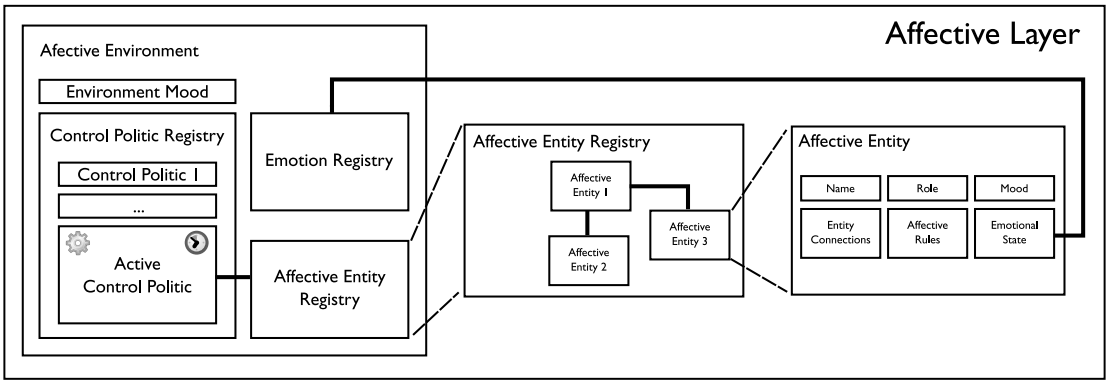


Figure 3.2: Affective layer internals

Affective Entities are the atomic affective component in the system. Each entity must be uniquely identifiable and maintain a semantic description of its role in the environment. Affective entities maintain their own **emotional state** consisting of a set of emotions, a mood value and a set of emotional rules. Additionally, they are able to establish connections between them.

Emotions are usually passive elements maintaining at least a "strength" value and when relevant, useful semantic information. There are no constraints on the type and scale of emotion or mood values, such decision is implementation dependent. To implement advanced emotional models, emotions can be provided with affective intelligence, e.g., the concept of "decay functions" introduced in the OCC model [67] of emotion. The emotions affective intelligence is a type of functional delegation.

Emotional Rules are attached to the emotional state of a specific affective entity, and they are the other delegation type. When the entity's emotional state changes, the attached emotional rules are evoked and eventually perform a set of actions over that same entity. Emotional rules were thought to implement constraints over the entity's emotional state, e.g., if the entity is happy the sadness value should be in its minimum. Emotional rules can be used to perform routine operations that otherwise should be explicitly requested by the affective application. Emotional rules can be registered globally, i.e., new affective entities have these rules automatically attached to their emotional state.

Entity connections consist of three mandatory elements: the unique identifier of the affective entity towards which the connection is established, a semantic qualifier denoting the character of the connection and a strength value. Connections are unidirectional, bidirectional connections are represented by two unidirectional connections, one in each affective entity.

3.2.2 Composition layer

The composition layer - figure 3.3 - supports the composition algorithms. Music composition is done through a composition pipeline, controlled by a **Composition Manager** that maintains a registry of available pipeline filters and composers. The **Composition Pipeline** consists of four types of elements: a data source, processing filters, a data sink, and pipes, the connection elements. Data flows in one direction, and each filter can only be connected to a previous element and a following element,

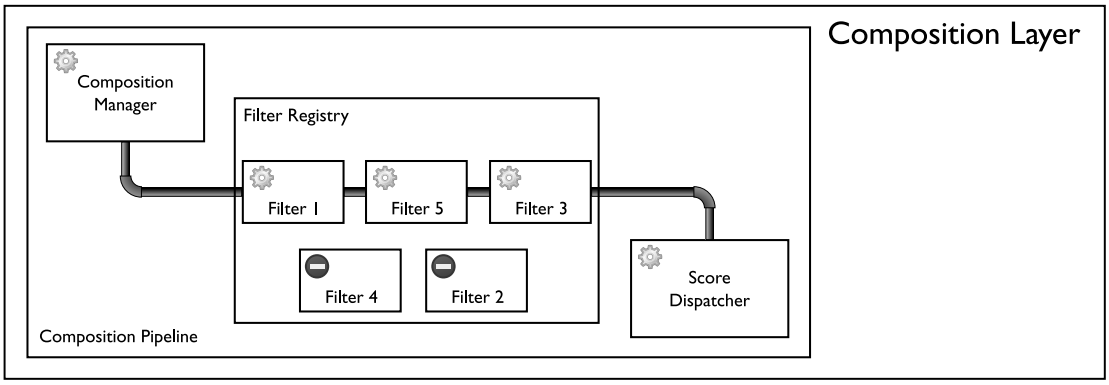


Figure 3.3: Composition layer internals

which means that it is not possible to split and re-assemble the data stream. The programmer is able to reconfigure the connections between the pipeline elements, but it should be careful to maintain a valid path betwee the data source and data sink. Inside the composition pipeline, data is transported in packets, but the type of transported data is undefined. Due to real-time constraints the composition pipeline is asynchronous, i.e., each processing stage – filter - has its own execution thread. The composition pipeline model is inspired in the Pipes and Filters design pattern.

The **Data Source** is the first element of the composition pipeline and receives information from the affective layer and passes it to the first filter in the pipeline. The **Data Sink** is the last element of the composition pipeline and receives information from the last filter in the pipeline and delivers it to the output layer.

Filters are the pipeline's construction blocks, its processing stages. They are active elements with their own execution thread. At each processing cycle, the filter reads data from its input port, processes it and writes the processed data on its output port. The input port must be connected to the output port of the previous filter or data source, while the output port should connect to the input port of the next filter or data sink. Filters can be activated or deactivated, being possible to change their state during run-time. A deactivated filter acts like a bridge between its input and output ports, i.e, it does not processes any data passing through it.

Pipes are the elements assembling the other pipeline components, they maintain an internal data buffer with variable size. Pipes provide non-blocking synchronised read/write operations, i.e., if it is not possible to perform the operation all acquired synchronisation resources are released and the calling thread should, eventually, try again later.

A **Composer** is a uniquely identifiable set of connected filters treated as a single entity. It is possible to register composer entities in the composition manager for later usage. The composition manager can load – and unload - composers during run-time, establishing the appropriate element connections automatically; however, at a given moment only one composer can be loaded. Composers are a convenient way to manage different composition algorithms.

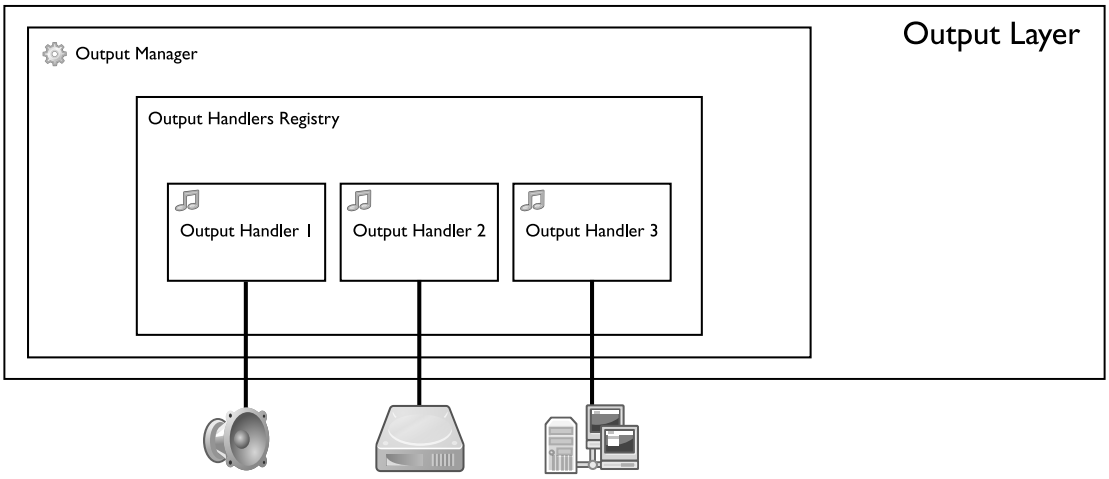


Figure 3.4: Output layer internals

3.2.3 Output layer

The output layer - 3.4 - is responsible for the output of composed music scores and maintains a registry of output handlers in the **Output Manager**. The most usual output are audio signals but the architecture imposes no limit on this. For example, it may be useful to write composed segments to a file – for later review in a music notation program - at the same time those same segments are played. An interesting set-up allowed by this architecture would be the direct connection to a music performance system.

Output Handlers encapsulate the output logic, data transformation and deliver to the desired output channel. It is possible to register more than one output handler, allowing for multiple simultaneous outputs. The only limitation to multiple output is performance. Due to scalability and real-time requirements, each handler must avoid processor intensive operations that can introduce output delays. For example, if an output handler must translate the internal representation of a music score to audio signals, the process should be fast enough so the music reaches the user's ears in time. Eventually, output handlers can be provided with internal buffers and their own execution thread, allowing asynchronous output. The implementation of such mechanism can be provided by an architecture implementation or delegated to the programmer. Individual handlers can be activated and deactivated.

3.3 Application Programming Interface

3.3.1 Affective API

Most of the interaction with the affective layer is done through the affective environment. Affective entities can be registered and unregistered in the affective environment, and it is possible to retrieve a specific entity from the registry. Emotion types must also be registered in the affective environment.

When a new affective entity is created its emotional state consists of the registered emotions and a mood value. The proposed architecture is neutral in what respects to emotion models, the programmer is free to implement the most appropriate models for his application. The affective API provides a base implementation for emotions that can be extended for that purpose. Emotional rules can be registered globally or in a specific affective entity, but regardless of the option, they are always registered in the affective environment. Global rules are active in every affective entity created before their registry, while specific rules are only active in the existent affective entities in which they were registered. Control politics are also registered directly in the affective environment. The affective API provides run-time switching between registered politics. Specific implementation of control politics are the programmer's responsibility. For that purpose an extensible base implementation is provided by the affective API. Connections between entities are registered in each specific entity. The normal procedure for registering a connection between two entities, is to obtain the origin entity through the affective environment and add the new connection. Finally, the affective API provides the necessary controls, to start, stop, pause and resume the affective layer.

3.3.2 Composition API

Interaction with the composition layer is done through the composition pipeline component. The composition pipeline provides the necessary functions to register and connect the pipeline's filters, including the data source and data sink. At a given time, only the connected filters are considered to be part of the composition pipeline. It is possible to re-design the connections between filters, i.e., reconfigure the composition pipeline during run-time. The composition API provides a base implementation for filters that should be used by the programmers to implement their specific filters. The composition API provides an alternative way for managing the composition pipeline. Composers are a set of filters along with connection information treated as a single entity, and it is possible to load and unload them during run-time. To define a new composer, the programmer encapsulates filter registration and connection operations in it, these operations are executed when the new composer is loaded. The composition API provides a model for music score representation. This internal representation is used in the composition algorithms to incrementally build a music segment, eventually translated in the output layer to the desired formats - usually audio. The programmer is free to use the provided representation, or to define its own. There are no special requirements over such representations. Finally, the composition API provides the necessary controls, to start, stop, pause and resume the composition pipeline.

3.3.3 Output API

Interaction with the output layer is done exclusively through the output manager. Output handlers must be registered before they start to process the musical segments delivered by the composition layer, an it is possible to temporarily deactivate each of this output handlers. The output API

provides a base implementation for output handlers, that should be extended by the programmer to define specific output handlers. If no output handlers are registered, or all of them are deactivated the system continues to function normally, but it simply outputs nothing. Finally, the output API provides the necessary controls, to start, stop, pause and resume the output manager.

3.4 Integration Driver

One of the main requirements of the proposed architecture is high integrability. This requires the definition of a model that should encompass a large set of different integration options and, at the same time, maintain the coherence from an architectural point of view, among different set-ups. The proposed integration model is based in the Driver concept. A **Driver** is a small program working between the affective application and the composition system. The driver receives information and control requests from the affective application and executes the adequate actions over the composition system through the proposed API. Communication with the affective application can be achieved through native code, if the driver is written in the same programming language than the affective application, using language wrappers or via network communication. To execute API operations over the composition system, the programmer can use native code or language wrappers. Any required data pre-processing such as, concept translation, should also be made in the integration driver, to keep these processing transparent to the affective application and the composition system.

3.5 Optional Functionality

3.5.1 Event notification

On most situations the data flow inside a composition system is unidirectional, i.e., data passes through a series of sequential processing stages until it is ready for output. However, it can be useful to distribute information in the opposite direction, through an event notification mechanism, so the previous processing stages can change their behaviour according to what is happening in subsequent stages. For example, it makes no sense to deliver more information for processing if the system is already having a hard time dealing with an excessive data flow. On the other side, such notification mechanism can be used to develop useful tools that can interact with the composition system such as, graphic user interfaces. The proposed notification mechanism consists of three main components: an **Event Manager** centralising event creation and notification, event listeners to capture and treat events and the events themselves.

Events consist of a type, an origin and optionally additional useful information for event treatment, and they are captured through event listeners that should be registered in the event manager. The event's life cycle can be resumed in three steps, a system component raises an event, the event arrives to the event manager and all listeners registered to receive that event type are notified, each notified listener receives the event and executes the adequate action, or eventually, ignores it. Two models of

listener registration can be followed, in an event basis, i.e., each event requires a different listener, or in a group of related events. In the later case is up to the programmer to decide which notifications to treat and which to ignore. Both models can coexist, but it is recommended that only one is provided for the sake of clarity and coherence.

3.5.2 Logging

Usually it is useful to analyse a the system's behaviour, after execution. In most situations this information is used for debugging purposes during the development process, but depending on the logs quality, it can provide valuable information over the system's internal operations.

The proposed log mechanism follows a centralized model, similar to the event notification mechanism. A **Log Manager** centralizes the **Log Requests** and notifies the set of previously registered loggers. Each **Logger** is responsible to decide if it is interested in the received messages or if they should be ignored. A **Log Message** has a type, usually describing the message's severity level or priority, an origin, a description and optionally a parameter providing additional information on the log event. A message log is a three step process, one of the system's components creates a log message and sends it to the log manager, registered loggers are notified, and finally the log message is treated or ignored.

Chapter 4

I-Sounds

"Music is everywhere. From the world emerges an hymn."

Victor Hugo

I-Sounds is a full implementation of the proposed architecture, and aims to be a complete and flexible tool for the development and integration of emotion-based composition algorithms. This implementation provides a basic run-time environment out-of the box, requiring no further customisation. For this purpose, concrete implementations of the required functional elements such as, control politics and output handlers, are available to the programmer that has no special requirements. However, this is rarely the case and while these concrete implementations are designed to be general and serve a generic scenario, they are more useful as implementation examples, the programmer will most probably need to provide its own implementations of such elements. For that reason, and in spite of providing a basic environment, the system is highly configurable and extensible to meet the programmer's needs, a great effort was put on providing the programmer with a complete and flexible programming interface. The I-Sounds API is fully compliant with the architecture's specification, but provides some additional tools that can assist the programmer wishing to use it as a development platform. I-Sounds is written in pure JAVA, ensuring the system's portability among different computational platforms. The current implementation requires version 6 of the JAVA SE development kit (SDK). The system is bundled in a single JAVA archive file for easy usage and distribution. Specific applications requiring advanced customization of the system's core, should use the source version instead, and eventually re-generate the JAVA archive file. Sources are available as Eclipse IDE projects that can be easily imported into other development environments if required.

4.1 Affective Layer

4.1.1 Emotion model

The proposed architecture does not specifies a particular emotion model. The I-Sounds implementation approach, is to provide the extensible concept of emotion that can adapt to meet special requirements, and by default, four emotions are provided: happiness, sadness, anger and fear. The specification postpones the decision on the scales of emotion intensity. In this implementation the scales were influenced by the widely used OCC model, emotion intensity is measured in a discrete scale, ranging from 0 to 10, and mood values - both for the affective entities and the environment – are measured in a discrete scale ranging from -10 to 10. It is important to note, that similarities with the OCC model end here, i.e., although the four base emotions are part of the 22 OCC categories and have a similar intensity scale, they are not a reduced implementation of the OCC model, namely, they do not implement concepts such as decay functions. This default set of emotions, does not, in any way, constrains the programmer in this respect, in fact, he is expected to implement his own emotions, appropriate, and perhaps only useful in his specific problem. Defining new emotions requires the programmer, to extend an abstract base class and provide the necessary code which can be arbitrary complex. For the new emotions to be used by the run-time environment, the programmer should register the new types, providing their full qualified class name to the relevant API methods.

4.1.2 Default control politic

Control politics are for the most part context dependent. However, the I-Sounds implementation provides a default control politic that averages the emotional state of all affective entities registered in the affective environment, and delivers the averaged emotional state to the composition layer. The emotion average, is a simple arithmetic calculus without any type of weighting. While this politic can be used as a fully functional politic, it is far more useful as an implementation example. Programmers implement specific control politics, by providing an implementation for the abstract control politic base class. In fact, the politics's thread structure and execution cycle are already provided, and the programmer is only required to implement the execution method. Custom politics, are registered in the affective environment by providing an instance of the class implementing the desired politic. Once registered, they can be used to control the affective layer behaviour and what is delivered to the composition layer.

4.2 Composition Layer

4.2.1 Composition pipeline and composers

As defined in the proposed architecture, the asynchronous composition pipeline is composed by four elements, the data source, the data sink, the pipeline's filters and pipes, the connection elements.

In I-Sounds the pipeline's data source is the interface between the affective and composition layers, affective data is delivered to the Composition Manager, the data source, which is responsible for feeding the first pipeline's filter. The pipeline's data sink is implemented in the Score Dispatcher, which pulls the composed sequences out of the pipeline, and delivers them to the Output Manager. Both the Composition Manager and Score Dispatcher have their own execution thread, started and stopped with the Composition Pipeline. The data source and data sink are therefore predefined, i.e., the programmer is not required to provide a specific implementation for those elements, but only the pipeline's filters or stages. Filters are registered in the Composition Pipeline, and can then be connected using the appropriate API methods. The I-Sounds API provides three connection methods, the first connects a filter to the data source, the second connects two filters and the third connects a filter to the data sink. Theoretically the data source and data sink can be directly connected, but that would not be of much interest. The connection elements of the pipeline, pipes, are buffered objects providing synchronised data access, with configurable internal buffer size through the I-Sounds configuration file. To build a composition algorithm the programmer implements a set of filters and then defines how they are connected.

The architecture's specification introduces the concept of Composer to facilitate the development and management of various, easy loadable, composition algorithms. When the programmer uses a framework such as I-Sounds, without the need to customise its internals, the development of a composition algorithm is a rather independent process, thus, it is convenient to develop them as independent projects sharing a common dependency on the development framework. The I-Sounds implementation simplifies the usage of Composers through helper functionality. Using the provided API, the programmer develops the algorithm stages and defines the connections between each filter, packing this information into a single archive - jar file. The I-Sounds run-time is then able to load the packed algorithm directly from the archive file into the composition layer. The system searches the archive for a manifest file that should pint the class implementing the composition algorithm and encapsulating its operations, i.e., filter registration and connection. The manifest file follows the Java manifest syntax, as well as the standard procedures for archive inclusion, it should be named MANIFEST.MF and located inside the archive under the META-INF directory. The two required arguments, Name and Class, are expected to provide the composer name and the fullqualified name of its implementing class, respectively.

4.2.2 Music score representation

The I-Sounds API provides a class hierarchy to represent music scores based in the Music XML specification. Music XML is a XML-based music notation file format, developed by Recordare LLC and deriving from older formats such as, MuseData and Humdrum. The main goal of Music XML, is to provide a complete language for music score representation while ensuring the maximum interchange flexibility, allowed by the usage of XML. The specification is currently in its second revision, dating from June 2007 and its adoption as increased since its first specification in 2004. The most wide

used notation programs are able to import and export music scores in Music XML, and since it is a public format, many open source notation projects also adopted Music XML as an interchange format. Appendix D features a music segment composed by AMADEUS, in traditional music notation and its corresponding Music XML representation. This internal music score representation, allows for a greater output flexibility. Using the provided marshalling and unmarshalling functionality, the composed music segments can easily be written in XML files, which can be later read by music notation programs. This enables an easy visual representation of the system's output, a valuable resource when the development of the composition algorithms is made in collaboration with musicians. In spite of providing this class hierarchy, the programmer is free to use an alternative - eventually simpler - internal representation for music scores. I-Sounds uses the Proxymusic Java class hierarchy for Music XML, distributed under a GNU General Public License.

4.3 Output Layer

The I-Sounds implementation provides a default output handler able to translate the music score internal representation into a multi-track MIDI sequence. The MIDI output handler supports most of Music XML features that are representable in MIDI, including different instruments and note velocity. By default, the handler uses the Java framework's MIDI sequencer and software synthesiser, but it is possible to use different MIDI devices, including hardware synthesisers, by changing the I-Sounds configuration file. To ensure improved sound quality, the usage of a better sound-bank, other than the one provided by default in the Java SDK, is recommended. Most applications will require the output of audio signals, thus, the MIDI output manager is adequate for most situations. However, if the programmer decides to use an alternative score internal representation this handler is not suitable, and in that case he must implement a custom handler to translate the alternative representation into the desired format, MIDI or other.

The I-Sounds implementation does not supports full parallel output by default, i.e., if more than one handler is registered output is done sequentially, i.e., a segment is only delivered to the next handler after the current handler finishes its operations. This maintains full compliance with the architecture's specification, which delegates the implementation of parallel output into specific implementations or the final programmer. For most usages, parallel output is not required, the common situation of audio and file output, works rather well in sequential mode. Nevertheless if the programmer needs full output parallelling support, it should provide buffered output handlers with independent processing threads.

4.4 Optional Functionalities

The proposed architecture refers two additional functionalities for the composition system, event notification and logging facilities. The I-Sounds implementation provides both. Custom listeners

are implemented by extending an abstract base class. There is one abstract class for each listener type that defines a number of abstract methods, one for each captured event. If the programmer is only interested in some events captured by the listener's type, then it should provide an empty implementation for the methods capturing the irrelevant events.

The logging sub-system is used by the system itself, but it is also accessible through the provided API, and programmers are able to use it for their own logs. Custom loggers extend from a logger abstract class, and should implement a single abstract method for message logging. Custom loggers are registered through the API in the Log Manager. Each individual logger maintains a set of handlers intended to manage different log resources, e.g., a set of files. The I-Sounds implementation provides a XML Logger that maintains six different logs, messages are filtered to be written in the approriate files. While it is not mandatory to write log messages in XML, this logger can be a reference for custom logger implementation.

4.5 System Configuration

While the I-Sounds implementation provides extension functionality for system customisation, its default behaviour can be configured. In particular, it is possible to configure the default control politic, set of emotions, composer and output handler, used upon initialisation of the run-time environment. The system's logging facilities are also configurable, although this is perhaps less relevant since the provided defaults fit most usages. System configuration is done through configuration files, that follow the Java's properties file syntax. There are two configuration files, one for core configuration and one for logging configuration. The core configuration file has the name, "I-Sounds.properties" and it is located in "*<project directory>/etc*", the logging configuration file is named "logging.properties" and it is located in the same project sub-directory. The default configurations and file syntax can be consulted in appendix E.

4.6 I-Shadows Integration

4.6.1 Interaction

The I-Shadows characters are affective agents with OCC minds. The OCC model of emotion [67] defines a set of 22 emotional categories; however, AMADEUS will only generate music from four of those categories: joy, distress, anger and fear, thus, only these four emotions are considered to integrate with I-Shadows. For each character - currently in scene - I-Shadows creates an affective entity in I-Sounds whose emotional state consists of those four emotions and a mood value. As the characters evolve in the story, the I-Sounds system is notified to update the entities emotional states as well. From the run-time environment side, a custom control politic monitors the emotional state of a particular affective entity and sends this information to the composition layer. The composed segments represent the emotional state of the currently monitored affective entity, e.g., Hero.

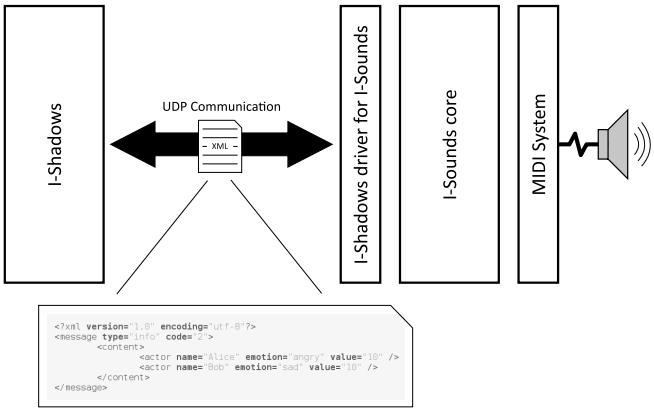


Figure 4.1: I-Shadows/I-Sounds integration

4.6.2 I-Shadows driver

The I-Shadows driver provides the glue between I-Sounds and I-Shadows, it serves two main purposes, the first is to provide a communication channel between I-Sounds and I-Shadows - which are implemented over different platforms - and the second is to hide the integration details from the two systems, i.e., for I-Sounds the driver is just an application interacting with it through its API, and for I-Shadows it is just an application receiving affective information through the network. The driver is written in Java, the same programming language of I-Sounds, thus, it uses native calls to the system's API to interact with the later. Communication is achieved through an XML-based message protocol - appendix C - using an UDP datagram stream. Two main reasons lead to the adoption of XML, it is a widely adopted solid standard supported in most programming environments with a rich set of manipulation tools such as, parsers and class domain translators, and on the other hand text based formats are easily encoded and decoded among different platforms. The network integration model allows both systems to run independently in separate machines, which can be important to run resource intensive. As a stateless unreliable transport protocol, UDP, has a low overhead when compared to its reliable companion, TCP, which makes it better suited for real-time applications. where data has a relevant time dimension. composition algorithms.

Chapter 5

AMADEUS

"Rhythm has something magic enough, to make us believe that the sublime can be ours."

Johann Goethe

AMADEUS is a composition algorithm implemented on top of the I-Sounds framework, able to compose short music segments expressive of certain emotions, those provided as the algorithm's input. AMADEUS is named after the eighteenth century composer, Wolfgang Amadeus Mozart.

From an affective point of view AMADEUS development was centred around a reduced set of emotions: happiness, sadness, anger and fear. The reader, may at this point, be wondering about the criterion leading to these basic set of four emotions. Happiness and sadness, are two popular choices in emotion-based composition, most of the studies reviewed in chapter 2 emphasise these two emotions. The main reason for such popularity, might be related with their bipolar character, i.e., happiness and sadness are conceptualised as bipolar emotions, and due to this fact they are easier to express and distinguish. Indeed, this same idea is sustained by Gabrielsson and Juslin [27, 50]. On the other hand, anger and fear are less studied, in part because they are harder to define and in part because they are harder to distinguish, exhibiting common expressive traces. However, this is exactly what makes them good candidates for study too. These emotions can effectively test the limits of music emotional expression, showing if music is able to unequivocally express these two emotions, or instead, their mapping in terms of musical factors is so blurring that musical segments tend to sound equally.

Music is the product of many different structural factors and the long list of section 2.3.1 is demonstrative of that. While all these factors have a potential contribution to the expression of musical emotions, the full understanding of that contribution is far from being complete. Even if such understanding was complete, the problem would remain far from being understood because emotion perception is also affected by factor interactions, i.e., the presence of a certain factor can change the

measured influence of other. In what respects to this matter, interactions are virtually endless and their impact on the perception of musical emotions remains unknown, at least for the most part. It is therefore important for the study of musical emotions to carefully isolate each individual factor in study, and keep others stable. Only by controlling the systematic variation in a factor while keeping others unchanged, it is possible to account each individual factor for the observed perception changes, and claim reliable results. Of course that this alone does not solves the problem of hardly controllable psychological factors that affect the listeners emotional perception; however, the researcher should at the best of his efforts that this influence is minimised. Building a composition algorithm that uses a relative high number of structural factors at once, is therefore a extraordinary hard task, even for a large passionate development team, thus, the wisest option is to adopt a constructivist approach, .i.e., systematically consider one factor at a time. While complexity will naturally emerge in such constructivist approach, as the number of considered factors increase, its emergence is gradual, which at least will provide the researcher with a better "problem tackling" ability.

These considerations imply some constraints over AMADEUS composed segments. The first of which, is the composition of short music segments, consisting of a single bar. One may ask if single bar segments are not to short for the listener to perceive the expected emotion. There are evidences, that musical emotions are quickly perceived by individuals within the same culture; studies by Peretz et al. [70] show that an adult individual - without special music knowledge - needs less than a quarter of a second to reliably distinguish the tone of a music excerpt as being happy or sad. A quarter of a second is just the enough time for a chord or a handful of notes. All segments are composed in a quaternary meter, notated by a 4/4 time signature, at the rate of 80 beats per minute. Melodically, only three notes: C,E and G, will be used to imply a C major or a C minor key in the segment.

A constructivist approach should focus, initially, on a reduced set of factors with broad structural influence, i.e, those providing the sketch filled with the artist's talent. Unfortunately, this seems to deep the problem even further by issuing another fundamental question of no simple answer, that serves as the introduction and motivation for the next section; what are those fundamental base factors?

5.1 Just in Time, a Theory of Rhythm and Meter

The *Just in Time* theory by Eduardo Lopes [56], was briefly introduced in section 2.4, and this is the adequate time to discuss it in more detail. Justice shall be made to the theory's name, because it uncovers evidences and traces of those pervasive and fundamental musical factors, that AMADEUS seeks. George Rochberg, refereed to music as "an intermediary link between the physical universe and ourselves". Lopes starts his work by commenting the following on Rochberg's idea;

> This is mainly because music relates physical vibrations (i.e. measurable properties of sound) with some kind of human form. The special feature of that human form is the creation of a temporal order without which sound could not be raised to the level of music.

It is undeniable that rhythm is one of the most important dimensions of music. As a percussionist, with his inherent deep understanding of rhythm, Lopes goes a step further on this idea and sustains that rhythm's influence is of paramount importance in music cognition, ans presents empirical evidences. Namely, he refers the works of Drake, Dowling and Palmer [19, 20], in which the perceptual significance of accent structures entirely derived from durational parameters, are compared with those deriving from pitch. In the words of Lopes, the results, "reveal a minor but systematic difference favouring the perceptual importance of duration-related parameters". However, theoretical formulations on music perception, have historically attributed a higher relevance to the pitch structure, to which rhythm is anchored. Opposing this traditional conceptions, Lopes proposes a rhythm-to-pitch approach for his theoretic formulations, i.e, one that models music perceptual qualities as emerging exclusively from durational parameters. The author frames this trend, in the contemporary Musicological shift towards the listener, which in his view, together with the emergence of music studies in cognitive science, lays the ground for a major change in the conceptions and importance given to music's durational parameters. To further explain his view, Lopes quotes the following by Dowling and Harwood [18], which is also very relevant in this work's context;

> The neglect of rhythm was especially unfortunate for the psychology of music because rhythm information is, if anything, more fundamental to music cognition than pitch information... We can recognise familiar tunes from their rhythmic patterns alone ... and when listeners are given an array of brief melodies in a multidimensional scaling task designed to find the important stimulus dimensions, rhythmic information tends to dominate pitch information on their judgements.

Lope's theory addresses the paradoxical secondary role given to music's durational parameters - at least in Western music theory - by developing, in the words of the author, "a systematic rhythm and meter model that is closely related to the perception of music". Lopes starts by analysing the perceptual behaviour of rhythm and meter in conjunction, then, he splits his analysis and considers each factor in isolation until he finally re-assembles them into a single "perceptual quality", to ultimately identify the perceptual qualities emerging from both factors. The result is an analytical tool to predict the perceptive qualities of different durational patterns, i.e., rhythms and rhythmic sequences.

Lopes work shares a common philosophical orientation with AMADEUS; he defends that music is more "democratic" on its surface than in higher levels, .i.e., more accessible to the common listener, barely captivated by deeper experiences, commenting on the perceptual qualities of music's surface;

> As opposed to pitch relationships, which show up at all levels of conventional music theory, durational relationships (rhythm and meter) are usually seen as confined to the musical surface, and thus as lacking critical value – as not, in short, a concern of the connoisseur: whereas it is my belief that it is exactly the perceptual qualities of the music surface, of which rhythm and meter are among the most important, that should be the concern of a music theory aiming to do justice to the listener.

The main goal of AMADEUS is to compose emotionally meaningful music, that can be used by affective systems to communicate with its user's. As such, and since the user's are in fact the listener's of AMADEUS, it is at the "democratic level" of music's surface - to paraphrase Lopes - that this goal is more likely to be achieved.

5.2 Affective Rhythm

Chapter 2 provided a brief empirical anthology concerning different structural factors and their influence on perceived emotions, including rhythm. From that discussion, it is clear that rhythm is among the least studied factors, and the existent research very superficial and general, not mentioning the terminological inconsistencies among different studies. Thus, AMADEUS required a solid theoretical foundation as the one presented above, that could not have been provided by such scarce research. Section 2.4 introduced the concepts of pulse salience and kinesis. These two concepts emerge from Lopes work as the two fundamental perceptual qualities of rhythm. AMADEUS takes on these two concepts, originally thought for music analysis, and applies them to affective composition. To do this, two main steps were given, the first was to develop appropriate property measures, while the theoretical background states that these qualities can be measured, it does not provides such objective measures, the second step was to map pulse salience and kinesis, using the developed measures, into regions or emotion clusters in a bi-dimensional model of emotion, inspired on Russel's circumplex model [79]. By combining this mapping information with the original Russell's model, hybrid maps of emotions and rhythm perceptive qualities can be built, and composition rules can be inferred and programmed in AMADEUS. This type of mapping approach was already explored, by Emery Schubert [85], but in a different context, the measurement and time series analysis of emotion in music.

5.2.1 The property of pulse salience

Salience quantification

Pulse salience was defined as the perceptive quality of rhythm, determining the relative "emphasis" of each pulse in a rhythmic sequence, i.e., some pulses are more prominent to the listener than others. This section proposes a quantification mechanism for pulse salience. As noted in section 2.4.1, pulse salience results from three components: metric position, agogic accentuation, and rhythmic cell accentuation. Thus, it can be modelled as sum of three individual components. Since it is not possible to account for the individual contribution of each component, i.e., to assess what components are more important, the following discussion assumes an equal contribution.

Metric posisition. Music theory, long acknowledges that some beats are stronger than others, i.e., pulses placed in these strong beats tend to sound more stable. These stable beats are very important in the perception of rhythm and meter, because they determine how a rhythmic sequence is organized by listeners. In a quaternary meter for instance, the first beat of a measure is considered as the

most stable, followed by the third, the second and the fourth, considered to be unstable. The relative stability of each metric position can be extended to a finer grain, .i.e, the internal beat organization follows the same structure. For quantification purposes, each possible metric position is assigned a successive number $n \in N$. The attributed values are inversely proportional to the positions relative stability, i.e., the strongest position receives the smallest value while the weakest receives the greatest value. Considering;

B the number of beats in each measure.

U the value filing a beat.

M the value of the shortest pulse present in the rhythmic segment.

The metric component is given by (5.1).

$$O(\omega) = BU - M\omega \tag{5.1}$$

Where;

ω is the pulse's metric position.

Agogic accentuation. Agogic accentuation depends exclusively on the pulse's duration, the longer the pulse the more stable it is perceived. Note however, that this is measured not in terms of the pulse's absolute duration in time, but in terms of his relative duration. In music, each pulse has a relative duration whose absolute value in time, depends on the piece's tempo, for this purpose, whole notes are considered to be the unit that can be divided to denote shorter pulses, e.g., the half note has half of the duration of the whole note, the quarter note as half the duration of an half note, and a quarter of the duration of a whole note, and so on. For quantification purposes, each note type, e.g., whole notes, half notes and eighth notes, are assigned a number $n \in N$. The asigned values should reflect the fractional character of each note type, e.g., if a quarter note as a value of 1000, an eighth note will have a value of 500 and a half note a value of 2000. The agogic accentuation component of a pulse's salience, is the value corresponding to it's type, in other words, the agogic accentuation component is given by the identity function (5.2).

$$P(\eta) = \eta \tag{5.2}$$

Where;

η is the pulse's type.

Rhythm cell accentuation. The overall salience of a pulse is affected by the preceding rhythmic context, shorter pulses have an accentuation effect over a pulse's stability, but if the preceding pulse is equal or greater than the one being quantified then the value of this component is equal to 0. For quantification purposes, pulses are grouped in rhythm cells, each spanning for the time equivalent to a beat. To quantify the accentuation value, only the preceding rhythm cell is considered; however, there is an exception when the quantified pulse is preceded by two, or more, equal and homogeneous rhythm cells, i.e., rhythm cells with the same number of pulses, consisting of only one type of pulse, e.g., sixteenths, in which case the individual rhythm cells are considered as single big cell. Considering;

M the value of the shorter pulse present in the rhythmic segment.

C a set of tuples (ζ, σ) representing the preceding rhythm cell - that can be in fact a group of equal homogeneous cells. For each tuple, ζ is the value of a pulse type and σ the number of pulses of that type in the rhythm cell.

R a function given by (5.3).

$$R(x) = \begin{cases} x & \text{if } x > 1 \\ 0 & \text{if } x \leq 1. \end{cases} \tag{5.3}$$

The rhythm cell accentuation component is then given by (5.4).

$$Q(\eta) = \sum_{(\zeta,\sigma)\in C} \left(\left\lceil R\left(\frac{\eta}{\zeta}\right) M\sigma \right\rceil \right) \tag{5.4}$$

Where;

η is the pulse's type.

Pulse salience general formula

Pulse salience is therefore, a first order polynomial function, given by (5.5a), or its extended form (5.5b), reflecting the fact of being the result of three different components with equal weight.

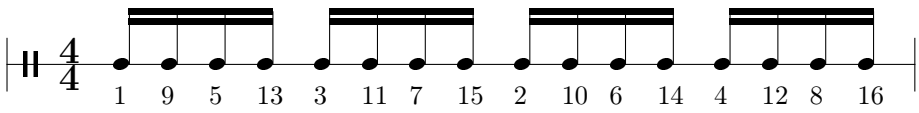
$$S(\eta, \omega) = O(\omega) + P(\eta) + Q(\eta) \tag{5.5a}$$

$$S(\eta, \omega) = (BU - M\omega) + \eta + \sum_{(\zeta,\sigma)\in C} \left(\left\lceil R\left(\frac{\eta}{\zeta}\right) M\sigma \right\rceil \right) \tag{5.5b}$$

Where;

η is the pulse's type.

ω is the pulse's metric position.



Score 5.1: The sixteenth metric positions in a quaternary meter

At this point, it is important to make some considerations about the meaning of pulse salience. Pulse salience is a relative measure, i.e., the meaning of a pulse's salience is only relevant inside the rhythmic context in which it was calculated, to compare the relative stability of the segment's pulses. It is not possible to say, that a pulse with a certain value of salience is more salient or stable than other pulse from other rhythmic context, that eventually has a lesser salience value. The absolute salience value is dependent on the values assigned to each metric position and pulse type, which might differ among different contexts. However, even if these values are carefully chosen to be equal, comparison between two different segments is not valid, e.g., it is possible for two pulses in different contexts to have the same salience value, but one of them be more stable because the rhythmic context preceding it determines so.

Practical application example

Expression (5.5b) is a general function for pulse salience quantification, i.e, it can be simplified for each metric context by substituting the expression's constants with appropriate values. The following example applies the general expression to a quaternary context notated with a 4/4 time signature.

In a quadruple meter each measure is divided into four beats - some authors consider quadruple meter as two duple meters tied together, thus, each measure has four basic metric points that can accommodate a quarter note each. However, it is not possible to compose music with quarter notes only, it is necessary to use other values, such as eighth and sixteenth notes. Although shorter notes may exist, they are more rare, hence, this example considers the shortest possible pulses to be sixteenth notes. Now, with shorter notes, each beat can accommodate more than one pulse, hence, more metrical points are needed within a beat. If each quarter note is substituted by four sixteenth notes – sixteenth notes have one quarter of the duration of quarter notes - the measure will have sixteen pulses, each with the duration of a sixteenth note, and since this example does not considers shorter pulses, sixteen is also the number of possible metric points. Score 5.1 shows a measure filled with sixteenth notes, one in each of the sixteen possible metric positions, the numbers below each pulse are the values assigned to each metric position, respecting the rules defined in 5.2.1.

This example considers pulse types, from the whole note to the sixteenth note, table 5.1 shows a possible assignment for pulse values, according to the rules defined in section 5.2.1.

With this information, it is possible to determine the value of the B, U and M constants; B, the number of beats in each measure is equal to 4, U, the value filling each beat is equal to 1000 (the value attributed to quarter notes), and M, the value of the shortest possible pulse in the rhythmic segment is equal to 250 (the value of sixteenth notes). Expression (5.6) can then be derived from (5.5b).

Table 5.1: Pulse types and their possible values

Pulse type	Notation	Value
Whole note	𝅝	4000
Half note	𝅗𝅥	2000
Quarter note	♩	1000
Eighth note	♪	500
Sixteenth note	𝅘𝅥𝅯	250

Table 5.2: Relation of rhythm and loudness with the basic set of emotions

	Happiness/Joy	Sadnes/Distress	Anger	Fear
Rhythm	Regular/Smooth Varied Flowing/Fluent	Firm	Complex	-
Loudness	Loud	Soft	Loud	Soft
Loudness Variation	Small	-	-	Large Rapid changes

$$
\begin{aligned}
S(\eta,\omega) &= (BU - M\omega) + \eta + \sum_{(\zeta,\sigma)\in R}\left(\left\lceil R\left(\frac{\eta}{\zeta}\right)M\sigma\right\rceil\right) \\
&= ((1000 * 4) - 250\omega) + \eta + \sum_{(\zeta,\sigma)\in R}\left(\left\lceil R\left(\frac{\eta}{\zeta}\right)250\sigma\right\rceil\right) \\
&= (4000 - 250\omega) + \eta + \sum_{(\zeta,\sigma)\in R}\left(\left\lceil R\left(\frac{\eta}{\zeta}\right)250\sigma\right\rceil\right)
\end{aligned}
\tag{5.6}
$$

Pulse salience and emotions

At the best of this work's research efforts, the relation between the property of salience, as well as kinesis, with emotion, was never directly considered. For that reason, the present work advances with its own hypothesis, based in existent research on other related factors. In what concerns to pulse salience, three structural factors in special were considered, rhythm in its generalist analysis, loudness and loudness variation. A summary of their qualitative relations with the basic set of four emotions is presented in table 5.2, built from the literature review in chapter 2.

Due to the relative character of the salience measure, it is also not possible to map absolute values of salience to specific emotions or categories, therefore, this mapping can be at most, a qualitative relation between pulse salience and emotion categories. Figure 5.1 shows the hypothesized mapping of pulse salience into a bi-dimensional model of valence and activation.

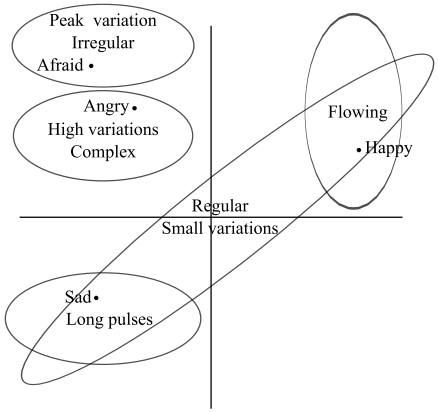


Figure 5.1: Proposed mapping of pulse salience into the bi-dimensional model

5.2.2 The property of kinesis

Kinesis refers to the perceptive quality of rhythm responsible for music motion perceived by listeners, and its primary source is meter, i.e., the different stability of each metric point. Placing pulses in unstable or week metric points contradicts the natural metric placement of pulses, rising tension levels as a consequence, which can resolved, by returning to a natural pulse positioning. The creation and release of tension has a direct relation with the the perception of musical motion. On the other side, kinesis, is also related with pulse density, i.e, using shorter pulses to fill the same time interval accentuates musical motion because shorter pulses are perceived as less stable.

For kinesis, a qualitative measure seems more suitable than a quantitative measure such the one proposed for pulse salience. Pulse salience is a property that can be quantified for each individual pulse, whose values can then be compared among each other. On the contrary, it does not makes sense to measure kinesis on a single pulse, but for a whole segment or sub-segments. A quantitative comparison would be artificial and less important.

Kinesis and emotions

Such as pulse salience, the property of kinesis was not mentioned in the significant and representative studies reviewed in chapter 2. Likewise, the present work proposes a mapping hypothesis for the rhythm quality of kinesis, partially based in the considerations made about tempo, summarised in table 5.3. Tempo can also influence the perception of musical motion, and thus important cues can be derived from its study. As a general rule, faster tempo implies more motion, while slower tempo implies a more relaxed experience. AMADEUS maintains a constant neutral tempo of 80 beats per minute, not to fast or to slow, but close to our heart beat. The control of kinetic potential, is achieved through the manipulation of pulse density and metric placement. A higher pulse density implies the usage of shorter, less stable pulses, increasing the kinetic potential of the composed sequence, e.g., a sequence consisting of many sixteenth notes elicits more kinesis than a sequence using more eighth notes. Figure 5.2 shows the hypothesized mapping of kinesis into the bi-dimensional model.

Table 5.3: Relation of Tempo with the basic set of emotions

	Happiness/Joy	Sadnes/Distress	Anger	Fear
Tempo	Fast	Slow	Fast	Fast

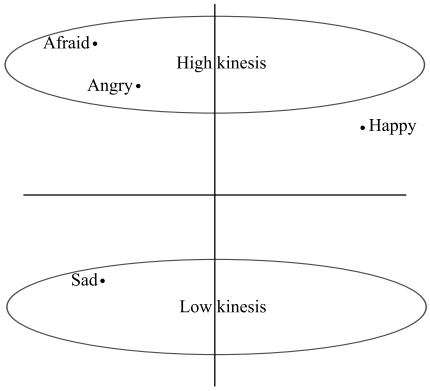


Figure 5.2: Proposed mapping of kinesis into the bi-dimensional model

5.3 Diatonic Modes

Although AMADEUS puts a special emphasis on rhythm, it uses the diatonic major and minor modes as an additional resource for emotion expression. Mode is one of the most consensual factors among researchers, i.e., there is an general agreement on how it influences emotion perception, extending to Musicology and among musicians. In fact, the ancient Greek philosophers were the first to discuss the connection between modes and emotions. The major mode is usually classified as happy and joyful, while the minor mode is usually classified as sad and melancholic.

The most important distinctive feature in a diatonic scale, determining its modes, is the interval formed by the scale's tonic and mediant, i.e., the fist and the third scale degrees or notes. If this interval consists of two whole tones - a major third interval - the scale is said to be in the major mode, on the contrary if the interval consists of a whole and a half tones - a minor third interval - the scale is said to be in the minor mode. However, for the complete perception of mode, at least a third note is necessary. The fifth scale degree of a diatonic scale - named dominant - is the most adequate for such purpose. The tonic, mediant and dominant scale degrees, form what is sometimes named of tonic triad, a central concept for tonal music. If played simultaneously as a chord, or sequentially, these notes can define the scale's tonal centre and mode. In what respects to melody and the diatonic modes, AMADEUS uses these three notes only. Once rhythm is composed, each pulse is assigned a pitch in ascending and descending sequence, repeating if necessary, e.g., for a sequence with four pulses and a C major context, the assignment order will be, C-E-G-E. To maintain a clearly distinctive major and minor contexts, AMADEUS, should start each single measure segment with a C and terminate with a E. This imposes an additional constraint on the composed segments, since there are 3 different pitches - C,E and G - each measure should have, 4, 8, 12 or 16 pulses so the above constraint can

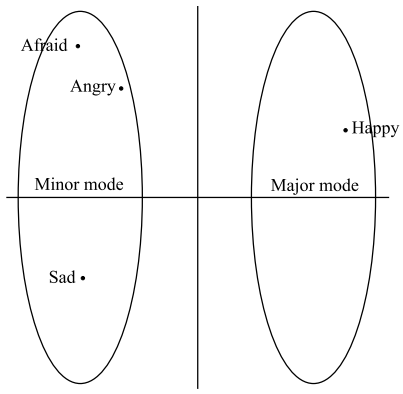


Figure 5.3: Proposed mapping of the diatonic modes into the bi-dimensional model

be respected. A different number of pulses would require pitch transitions between measures different from E-C, which affects the perception of mode. Based in the discussion of chapter 2, figure 5.3 shows the mapping of the diatonic major and minor modes, into the bidimensional model of emotion.

5.4 Implementation

The AMADEUS composition algorithm consists of three stages - filters in the I-Sounds terminology - the Base Filter responsible to construct a base representation of the score that is sequentially updated during the next two pipeline stages, the Rhythm Filter responsible for laying out the segment's rhythms, and the Mode Filter responsible for the melodic aspects of the segment. These three stages are connected in that same order. This brief section discusses relevant implementation aspects of AMADEUS filters, specially of the Rhythm Filter.

5.4.1 Base filter

AMADEUS base filter, does not provides any composition functionality, i unique function is to build a base score, using the provided class hierarchy. The base score, consists of one staff decorated with a G clef, and a 4/4 time signature, while tempo is kept at a constant rate of 80 beats per minute. Additionally, this filter defines the instrument that should play the short composed sequences - the familiar timbre of a piano in the case of AMADEUS.

5.4.2 Rhythm filter

The Rhythm Filter is the most complex stage of AMADEUS, and it is responsible to fill each of the segment's beats with suitable rhythmic constructs, i.e., a set of one or more notes or rests. AMADEUS uses a rule-based system for rhythm composition in the form of a CSP, the acronym for Constraint Satisfaction Problem, a problem formalisation used in the Artificial Intelligence field.

Table 5.4: AMADEUS rhythmic pallete

Single note	Two notes	Three notes	Four notes
♩	♫	♫♬	
	♫.	♬♬	♬♬♬♬
𝄾♪	♬.		
	𝄾♬	♬♬	

Constraint Satisfaction Problems

CSP's are an optimized way to solve some types of problems, that would be less efficiently solved by general purpose search techniques. A CSP problem consists of a set of variables, $X_1, X_2, ..., X_n$, and a set of constraints, $C_1, C_2, ..., C_n$. Each variable X_i has a non-empty domain D_i of assignable values, and each constraint C_i imposes restrictions on the legal combinations of values for a set of variables. A problem state is defined as a set of assigned variables, and if the set holds all X_i variables, it is said to be complete. Any assignment that does not violates any of the C_i constraints is said to be consistent or a legal assignment. A problem solution, is then defined as a consistent complete state, i.e, a state in which all variables are assigned and none of the assignments violate any of the problem constraints. CSP's are treated with greater detail in the influential book, "Artificial Intelligence: A Modern Approach", by Stuart Russel and Peter Norvig [80]. AMADEUS search algorithm is inspired in Russel's and Norvig's pseudo-code, provided in the same book.

Rhythm composition CSP formalisation

Such as a real composer, the basic unit for rhythm composition in AMADEUS are rhythm cells. AMADEUS considers four variables, one for each measure's beat, whose domain of assignable values is a sub-set of AMADEUS rhythm pallet, given by table 5.4. The first beat in each measure must always be a quarter note, thus, the first variable's domain consists of a quarter note (♩) only. The last beat in a measure must always end with a sixteenth note, thus, the domain of the last variable is a sub-set of ♫., 𝄾♬, ♬♬, ♬♬ and ♬♬♬♬, except for sad measures that can end with a ♩. For each emotion, AMADEUS defines a set of salience intervals, one for each metric position, that should be respected by the pulses filling each position. This constitutes the primary control instrument over AMADEUS output. These intervals are not hard-coded, but defined in a configuration file, so they can easily be tuned without recompiling AMADEUS. The first evaluation test presented in chapter 6, is intended to provide parametrisation data for this intervals, by evaluating a series of candidate values. Additionally, AMADEUS should ensure that each segment is composed by 4, 8, 12 notes, so their pitches can be set appropriately. The CSP formalisation can then be summarised in table 5.5.

Table 5.5: CSP formalisation for rhythm composition

Variables	
	Domain values
First beat	♩
Second beat	The whole rhythm pallete as shown in table 5.4
Third beat	The whole rhythm pallete as shown in table 5.4
Fourth beat	(♩.♫ , 𝄾♫ , ♫♫ , ♫♫ , ♬♬) and (♩) in sad sequences
Constraints	
1. Each pulse's salience must respect the interval defined for its metric position	
2. Each composed segment should have 4, 8 or 12 pulses	

5.4.3 Mode filter

Once the rhythmic component is composed, the score is passed to the mode filter, where each pulse is assigned a pitch. As previously referred, AMADEUS uses a C context and only three pitches, C, E and G for major mode, or C, E flat and G for minor mode. Pitches are assigned sequentially, the assignment starts with a C and reverses the order when C or G are reached until all pulses are already assigned, e.g., a sequence with four pulses will be assigned, C-E-G-E, a sequence with eight pulses will be assigned, C-E-G-E-C-E-G-E and a sequence with twelve pulses will be assigned, C-E-G-E-C-E-G-E-C-E-G-E.

Chapter 6

Evaluation

"Music before anything."

Paul Verlaine

This chapter presents the empirical evaluation of the I-Sounds framework and the AMADEUS composition algorithm. The first experiment is a purely auditory test with two main objectives, to assess the effectiveness of AMADEUS composed segments in the expression of certain emotions, and to provide feedback data used to refine the algorithm parametrisation for the other two experiments. The second and third experiments were performed in integration with I-Shadows [4], an Interactive-Drama application. Both experiments combine the visual and auditive channels to assess the improvement in the recognition of the drama characters emotions, and the overall contribution of music to the intelligibility of the performed narratives, respectively. Each of the next three sections, present the methodological procedures, results and discussion for each of the three experiments.

6.1 AMADEUS calibration

AMADEUS uses a set of intervals to constraint the allowed salience value for each metric position - see section 5.4.2. In this first experiment, AMADEUS was parametrised with four interval sets, one for each expressed emotion: happiness, sadness, anger and fear. These candidate intervals are based in the author's empirical music knowledge. The experiment results are expected to provide useful data that can be used to further refine these candidate set of intervals, but also a measure of the expressive qualities of rhythm and mode. In this experiment two demographic factors are considered, the participant's genre and his music knowledge. The objective is to assess impact of these factors in the perception of musical emotions, therefore, this experiment addresses three fundamental research questions;

1. Are the proposed interval sets adequate to the expression of the respective emotion?
2. What is the expressive potential of rhythm and mode, in the expression of the same four emotions?
3. Does gender and music knowledge, affects the perception of musical emotions?

The hypothesised salience intervals are summarised in table G.1. Gender and music knowledge are expected to not affect substantially the perceived emotions. No hypothesis is advanced for the second research question.

6.1.1 Methodology

Design

This experiment consists of a single independent variable, the segment to be classified by the participant. Each segment is nominally classified as happy, sad, angry or frightening. The experiment consisted in the evaluation of 3 segments for each emotion for a total of 12 segments, listed in appendix F. All segments were composed in quaternary meter, denoted by a 4/4 time signature and played with a constant tempo of 80 beats per minute. Each segment varies in the rhythmic constructs and mode, C major or C minor. All other compositional and performance factors were, at the best of the experimenters efforts, kept stable. Six dependent variables were measured; gender, the participant's music knowledge, and the expressiveness of happiness, sadness, anger and fear, for each segment. The gender variable is obviously nominal, with the domain of male and female. Music knowledge, is also a nominal value, with two possible values, yes or no, denoting the answer given by the participant, when asked if it has some music theory knowledge and/or is able to play an instrument. To measure the other four variables a five point likert-scale was used, each of the scale's points is labelled with the following terms: disagree (lowest scale point), somewhat disagree, neutral, somewhat agree and agree (highest scale point). For statistical correlational treatment, these variables were considered to be ordinal. This experimental test follows a repeated-measures design, i.e., each participant evaluates all of the 12 segments, 3 per emotion.

Participants

A total of 68 volunteers, 53 males and 15 females, classified all of the twelve segments. From the total number of participants, 37 have some music knowledge about music theory and/or play an instrument, while 31 do not. No other demographic parameter was recorded for the population sample; however, the questionnaire was delivered among academic institutions where the typical minimum age is 18.

Apparatus

The test consists in a single web-based questionnaire with fourteen questions in total. The questionnaire was distributed electronically by email among a number of different research groups, some of

them with no direct connection with affective computing or computation at all. Music related research groups were not used in this test.

Procedure

Each participant was provided with a brief description of I-Sounds, AMADEUS and question types before starting the test. Special indications to take the test were also provided; the anonymity of the test, the exclusive usage of data for research purposes, the need to answer instinctively, the non existence of correct or wrong answers, and the elimination of any noise source that could influence the hearing of each segment. The participants were characterised, and grouped, by the first two questions, gender and musical knowledge. In the remaining 12 questions, each participant listened a short music segment of 9 seconds consisting of one single bar repeated three times, and classified it according to its expressiveness of happiness, sadness, anger and fear. The participants were allowed to listen each segment the number of times they considered adequate. At the exception of the two first questions, each of the twelve segments were presented randomly.

6.1.2 Results

Descriptive statistics

Histogram 6.1 shows that, participants agreed that happy segments where expressive of happiness, and disagreed that the other three types of segments were also expressive of happiness. This shows a clear matching between the happy segments and their perception as happy, i.e., the participants were clearly able to distinguish the happy segments from the others. Histogram 6.2 shows that sad segments were perceived as sad, although a significant portion of the participants agreed that they could also be interpreted as angry or frightening. Nevertheless, they did disagree that sad segments were expressive of happiness. Histograms 6.3 and 6.4, are similar, the angry and frightening segments were not perceived as expressive of anger and fear respectively. The histograms reveal low levels of agreement when the participants where asked if those segments were expressive of the respective emotion. It is also clear that angry and frightening segments were consistently classified with moderate levels of agreement for all four variables, suggesting that most participants were confused and had many doubts.

These histograms summarise the evaluation of each of the 12 tested segments grouped by their intended emotion; however, the analysis of each individual segment can reveal what segments were more representative of each emotion, and if the results were homogeneous among the 3 segments of each emotion. Histogram 6.5 does not reveals additional information confirming that participants were able to easily distinguish the happy segments from the other 9 segments. Histogram 6.6 reveals heterogeneous results among the three sad segments. From its analysis it is clear that the first sad segment F.4 was undoubtedly classified as sad, while the second generated more doubts. Histograms 6.7 and 6.8, confirm the tendency in the moderate classification of the angry and frightening segment.

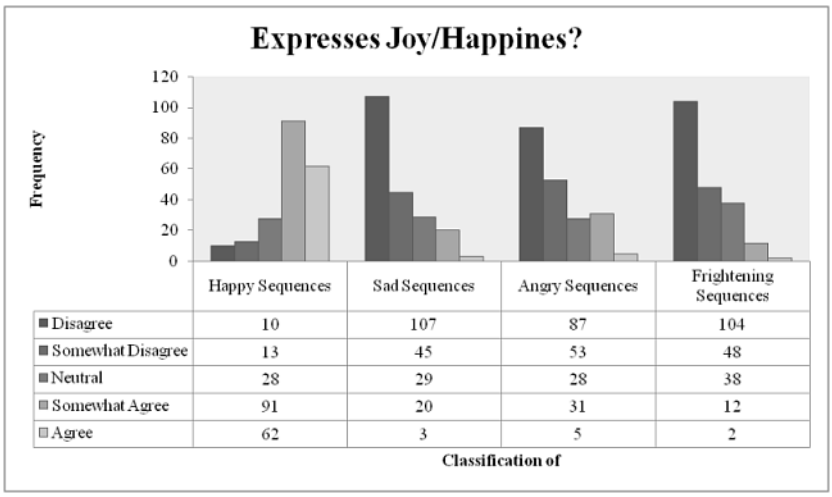


Figure 6.1: Ratings for the expression of happiness, grouped by the segment's emotion

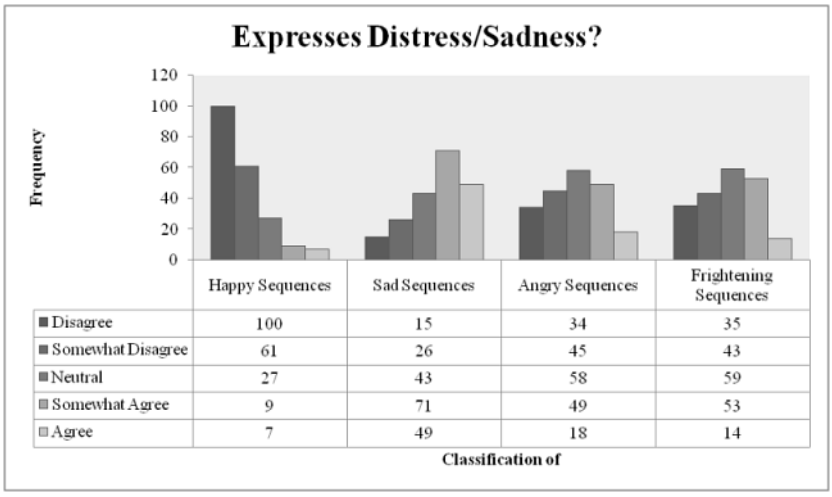


Figure 6.2: Ratings for the expression of sadness, grouped by the segment's emotion

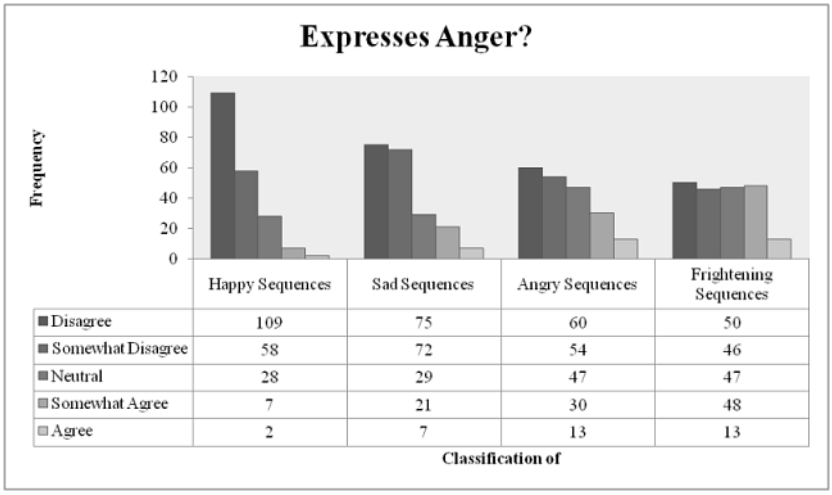


Figure 6.3: Ratings for the expression of anger, grouped by the segment's emotion

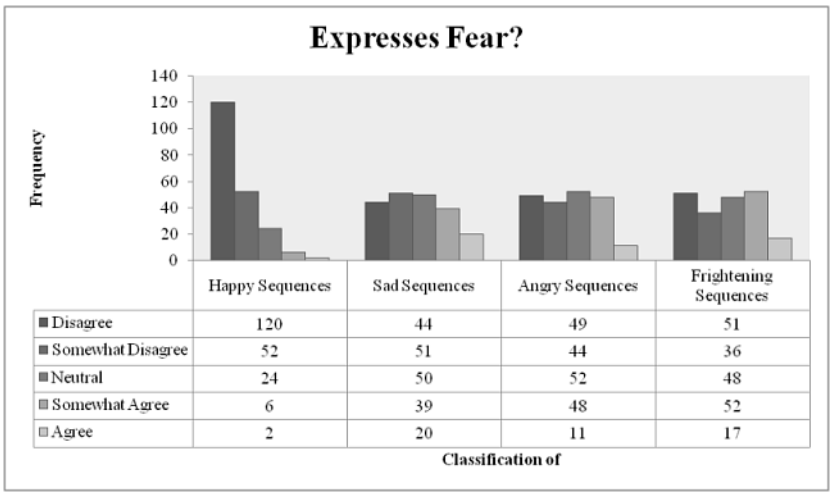


Figure 6.4: Ratings for the expression of fear, grouped by the segment's emotion

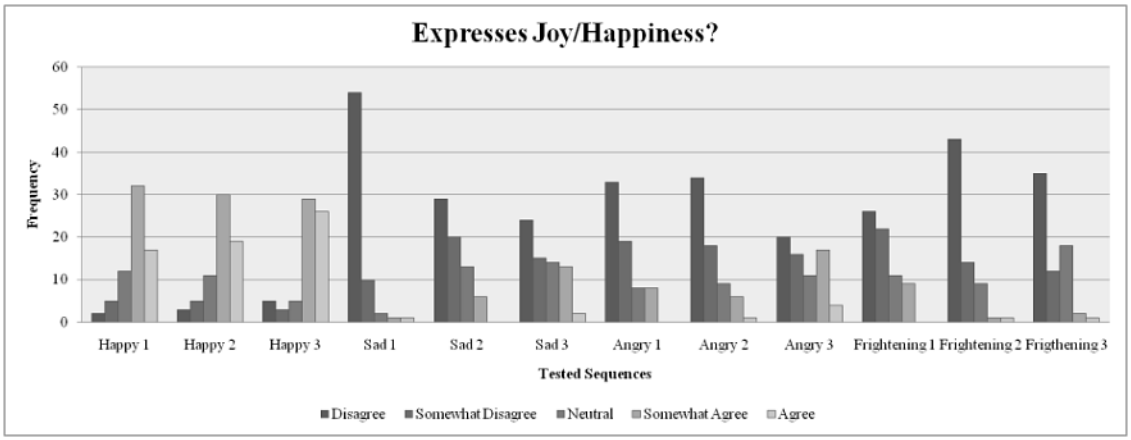


Figure 6.5: Ratings for the expression of happiness by segment

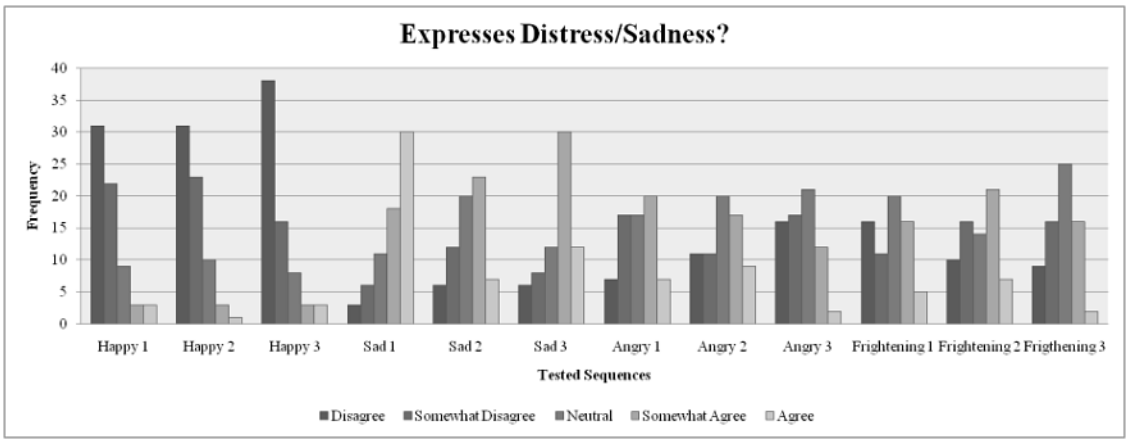


Figure 6.6: Ratings for the expression of sadness by segment

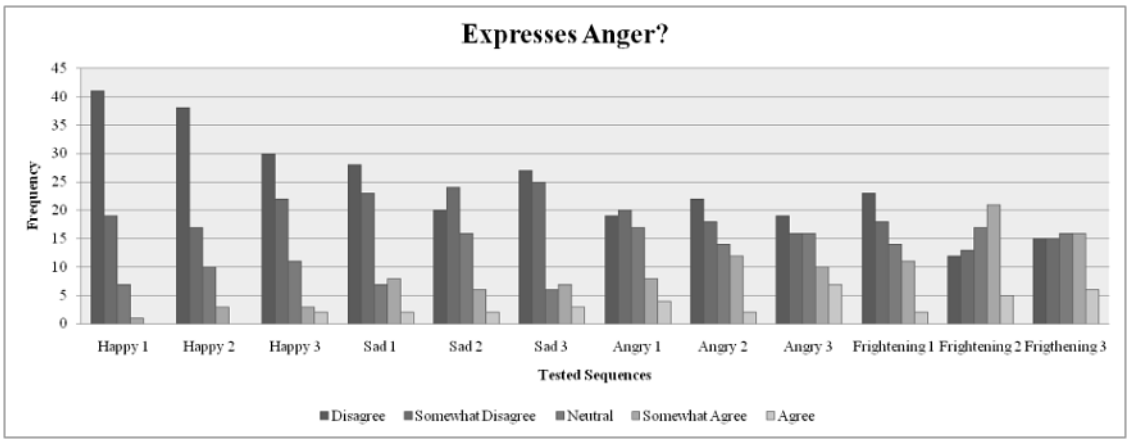


Figure 6.7: Ratings for the expression of anger by segment

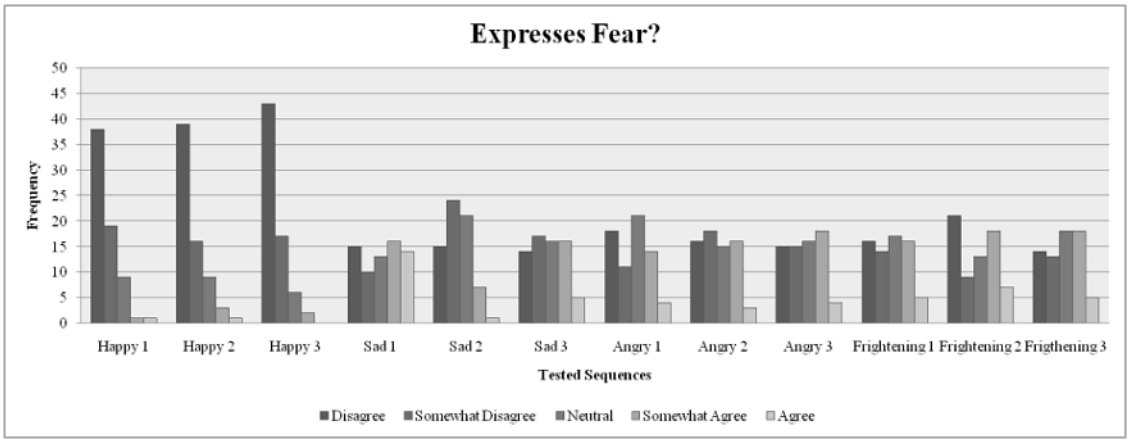


Figure 6.8: Ratings for the expression of fear by segment

Table 6.1: Spearman's rho correlation coefficients for the classification variables and the segments intended emotions

	Seq. Emotion	Exp. Happiness	Exp. Sadness	Exp. Anger	Exp. Fear
Seq. Emotion	1.0	-0.459	0.207	0.295	0.297
Exp. Happiness	-0.459	1.0	-0.433	-0.227	-0.398
Exp. Sadness	0.207	-0.433	1.0	0.168	0.281
Exp. Anger	0.295	-0.227	0.168	1.0	0.370
Exp. Fear	0.297	-0.398	0.281	0.370	1.0

Bivariate correlations

This section presents a correlational analysis, aiming to identify a correlation between the segments intended emotions and the attributed classifications. Instead of the traditional Pearson's correlational measure and due the the data's ordinal character, the following discussion uses the Spearman's rho correlational measure. No data operations were performed to remove outliers, in part, because Spearman's rho is less affected by outliers, and in part, because the effect of outliers in likert-scales is not very notorious due to the short range of classification values, 5 in this case.

Table 6.1 presents the correlation between the segments intended emotion and the participants classifications. While the perception of happy segments seems to be the only significant correlation with the segment's intended emotion, the low correlation in the perception of sad segments can be explained by the effect of the third, but mainly of the second sad segments. The results for the expression of anger and fear reflect the hard classification of the respective segments. An interesting exercise is to search for correlation between the perception of the four emotions. The expression of happiness and sadness have a considerable negative correlation coefficient, suggesting that segments highly classified as expressive of happiness tend to be classified low in the expression of sadness. The perception of happiness is in fact negatively correlated with the other three emotions, i.e., happy segments are clearly distinguished from the other 9. This suggests that happiness was the most easily recognisable emotion. Finally, the correlation coefficient between anger and fear, suggests a tight connection between the perception of these two emotions which was already suggested by the descriptive analysis.

Differences analysis

In this first experiment the participants were characterised by their gender and music knowledge. Mann-Whitney U tests were performed over the four dependent variables, concerning this two demographic factors, to identify significant differences in the participants classifications. The Mann-Whitney coefficients and respective significance, for each dependent variable, according to gender and music knowledge are shown in tables 6.2 and 6.3, respectively. In what respects to gender, the 2-tailed significance levels registered for all four classification variables are all significantly above the p-value

Table 6.2: Mann-Whitney U differences test, grouping by gender

Grouping variable: Gender				
	Exp. Happiness	Exp. Sadness	Exp. Anger	Exp. Fear
Mann-Whitney U	54471.500	54269.500	54032.500	57025.500
Asymp. Sig. (2-tailed)	0.303	0.276	0.232	0.937

Table 6.3: Mann-Whitney U differences test, grouping by music knowledge

Grouping variable: Music Knowledge				
	Exp. Happiness	Exp. Sadness	Exp. Anger	Exp. Fear
Mann-Whitney U	77797.000	77149.000	81406.500	82061.000
Asymp. Sig. (2-tailed)	0.138	0.097	0.715	0.872

of 0.05, which means that there is enough confidence to accept the null hypothesis, i.e., there are no differences in the segments classification that can be accounted to the gender of the participant. The same is verified to the factor of music knowledge, all significance values are greater than 0.05, suggesting that there are not significant differences in the participants ratings that can be accounted for that factor.

6.1.3 Discussion

The experiment results are very encouraging for the expression of happiness, the participants were able to clearly distinguish the happy segments from the others, and the correlational analysis suggests that happiness is the easiest emotion to communicate through rhythm and mode.

In what respects to the perception of sad segments, the results suggest that a refinement of the salience intervals constraining the composition of those segments might be required. The first sad segment F.4 was the only among the three sad segments to be clearly classified as sad, which makes it a good nominal reference for the refinement of salience intervals. The same segment received low classifications for the expression of happiness, a symptom suggestive of a bipolar relation between happiness and sadness, which is also supported by the negative correlational coefficient for the perception of both emotions. Indeed, this conclusion is in accordance with Russell's model of emotion, and supports the proposed mapping of mode and rhythm to the regions defined in figure 5.3 and 5.1.

On the contrary, the classification results for angry and frightening segments suggest a high degree of doubt and confusion. The participants were unable to clearly identify the intended emotion, and these segments were said to be expressive of anger, fear and sadness, at the same time. While this confusion might suggest incorrect salience intervals for both emotions, it also suggests that they are very similar. Indeed, Russel's model supports this observation representing anger and fear very closely in the circumplex model - see figure 2.1. Therefore, one can suggest that rhythm, by itself, is not

sufficient to represent such short distances - in valence and activation - between anger and fear, since minor mode was used the segments expressive of both emotions. To differentiate between such close emotions a third factor varying between two extremes is needed. An hypothetical candidate is the usage of lower and higher pitch registries, in relation to the central octave.

The differences analysis, were conclusive in what respects to classification differences between genders. The perception of emotion in music is not significantly dependent on this factor. Similarly, music knowledge seems to be an irrelevant factor; however, the significance value for the Mann-Whitney U test, registered for the expression of sadness - see table 6.3 - is very close to the p-value. In fact, if the one-tailed probability was considered, that value would be less than 0.05 and the null hypothesis would have been rejected, i.e., the existence of a difference, in the classification of sadness expressively, would be statistically significant. Thus, and although the statistical result - for the 2 tailed probability - is sufficient to accept the null hypothesis, i.e., the existence of a real difference is not to be totally excluded.

6.2 Emotion Recognition

One of the objectives of AMADEUS integration in the I-Shadows system was to enlarge the affective bandwidth between the system and the users. The recognition of the characters emotions, i.e., the ability of the user to recognise the emotion exhibited by a particular character at a given time, one of the most important measures of that affective bandwidth. This experiment was designed to measure the success rate in the recognition of the character's emotions in two different experimental conditions, with and without music accompaniment. The recognition rate was measured for the same four emotions considered in the first test: happiness, sadness, anger and fear. A single research question is proposed for this experiment;

1. What is the impact of music in the recognition of the characters emotions?

Music is expected to contribute positively for emotion recognition, i.e., an improvement in the recognition rate is expected for all emotions.

6.2.1 Methodology

Design

This experiment measured a single dependent variable, the recognition of the character's emotions. The variable can assume four categorical (nominal) values: happy, sad, angry and frightened. The experiment had a single independent variable, the presence or absence of music accompaniment in the presented clips. This experiment follows an independent measures design.

Participants

There was a total of 24 Portuguese boys and girls, in the scholar age between 8 and 9 years old. No further demographic factor was measured. All 24 participants attend the same school.

Apparatus

A single computer connected to a video projector, projected all video-clips. There were two experimenters with different tasks, one would interact with the young participants, explaining the objective of the test and helping with the interpretation of the questions while the other would remain behind the audience to control the projection of each video-clip.

Procedure

This experiment consisted in the projection of four video-clips, two times each, with and without music accompaniment, to two different test groups. Each clip depicts a character, exhibiting a certain motion pattern expressive of four possible emotions, happiness, sadness, anger and fear. After the projection of each clip, the participants were asked to immediately evaluate the characters emotion, using the evaluation matrix provided in appendix H.1. The projection of each clip could be repeated if the participants asked for. For logistical reasons, the 24 participants were divided in four groups of six children. These groups were arranged by their teachers, without any other criteria than the children's's regular scholar activities. Each of the four groups took the test in turn, and children were told to not comment the test with their colleagues, that were waiting for their turn. The first two groups assisted the video clips without music accompaniment, while the other two watched the clips with music accompaniment.

6.2.2 Results

Table 6.4 shows the character's emotion evaluation frequencies for each clip, in the condition of no music accompaniment, the relevant descriptive statistics and the results obtained for the Chi-square "Goddness of Fit" test, comparing the obtained and random distributions. Similarly table 6.5 shows the same statistic measures for the condition of music accompaniment.

The standard deviation registered for all clips in both conditions, are indicative of low statistical dispersion, suggesting an evaluation consensus among the participants. The 2-tailed p-values obtained for the chi-square test, are all less than 0,0001 except for the frightening clip in the condition of no music accompaniment, therefore, the null hypothesis that the collected data follows a random distribution, can be rejected in both conditions. There is a considerable increase in the successful recognition of happiness and sadness when music was added to the respective clips, and while a small increase was registered in the recognition of fear, a 12% decrease was registered in the recognition of anger. Of particular note is the 100% success rate registered for happiness, when AMADEUS happy compositions accompanied the happy clip.

Table 6.4: Results for the experimental condition without no music accompaniment

Clips		**Happy**	**Sad**	**Anger**	**Frightening**
Frequencies	Expresses Happiness	19	1	0	0
	Expresses Sadness	2	16	0	1
	Expresses Anger	3	1	15	12
	Expresses Fear	0	6	9	11
Descriptive statistics	Sum	24	24	24	24
	Median	1	2	3	3
	Mode	1	2	3	3
	Std. Deviation	0,7	0,93	0,49	0,58
Chi-Squared test	Chi-square	38,33	25,00	27,00	20,33
	Degrees of Freedom	3	3	3	3
	P-value (2-tailed)	$<$ 0,0001	$<$ 0,0001	$<$ 0,0001	0,0003
Recognition success rate		**79,17%**	**66,67%**	**62,50%**	**45,83%**

Table 6.5: Results for the experimental condition with music accompaniment

Clips		**Happy**	**Sad**	**Anger**	**Frightening**
Frequencies	Expresses Happiness	24	0	0	0
	Expresses Sadness	0	22	1	0
	Expresses Anger	0	7	8	12
	Expresses Fear	0	1	9	12
Descriptive statistics	Sum	24	24	24	24
	Median	1	2	3	3,5
	Mode	1	2	3	3
	Std. Deviation	0	0,45	0,56	0,51
Chi-Squared test	Chi-square	72,00	57,00	22,33	24,00
	Degrees of Freedom	3	3	3	3
	P-value (2-tailed)	$<$ 0,0001	$<$ 0,0001	0,0001	$<$ 0,0001
Recognition success rate		**100,00%**	**91,67%**	**58,33%**	**50,00%**

6.2.3 Discussion

This experiment brought mixed results in terms of emotion recognition improvement. The most encouraging results were recorded for happiness, indeed a 100% recognition rate was achieved when the various clips were projected with music accompaniment representing an increase of almost 21% over the condition of no music accompaniment, undoubtedly, a significant improvement. Although the recognition of sadness did not achieved the same success rate, its improvement was also clear and even greater in absolute value, the recognition rate registered with no music accompaniment was improved in 30% when music was introduced in the clips. The results for the other two emotions are less clear, in absolute terms the recognition of anger suffered a slight decrease from the condition with no music accompaniment, 62,50%, to the condition of music accompaniment, 58,33%. This odd result contrasts the improvement tendency for the other three emotions. Such a small difference might well be circumstantial, nevertheless, it can not be considered an improvement. In the recognition of fear, the difference registered between the two conditions is similar but positive. This difference can also be circumstantial, but on the contrary, it represents an improvement.

The results achieved for anger and fear, more than conclusive, are indicative of the need of further refinement on AMADEUS salience parametrisation values, which was already concluded in the previous test presented in section 6.1. It is also important to note, that the results could have been slightly different if other anger and frightening AMADEUS composed segments would have been tested, i.e., the choice of the segments could eventually had been better. An experiment to test all composed segments for each emotion, would require a very large sample and a very large experiment, incompatible with the logistic and timing restrictions.

In spite of these mixed results, that call for further refinement in AMADEUS, the usage of music can not be said to have decreased the ability of the user's to successfully recognise the character's emotion, but significant improvements were achieved for happiness and sadness, thus, it can be concluded that this results support, at a large extent, the experiment's hypothesis, of a positive effect of music in the recognition of the characters emotions.

Finally, an interesting comparison exercise between the original results achieved in the development of I-Shadows, which were purely visual, can be done, without forgetting that different experimental conditions such as the participants age, can influence the results. Nevertheless, and mainly for reference purposes table 6.6 compares the results achieved in the condition of music accompaniment with the ones originally registered [4] in the development of I-Shadows, denoting an overall and significant improvement in the recognition of happiness, sadness and anger, while the recognition rate of fear remained stable.

6.3 Story Intelligibility

The measurement of the emotion recognition rate with isolated clips featuring a single character provides a good experimental control, and it is an adequate approach to quantitatively assess music's

Table 6.6: Success rates in the recognition of the characters emotions

	I-Shadows	**Condition**
	original recognition rates	**with music accompaniment**
Happiness	70%	100%
Sadness	30%	92%
Anger	40%	58%
Fear	50%	50%

Table 6.7: Story intelligibility evaluation questionaire

Question 1	What were the boy and girl doing, at the beginning of the story?
Question 2	How the boy and girl felt, at the beginning of the story?
Question 3	How the boy and girl felt when the dragon appeared?
Question 4	How the fairy felt when she entered the story?
Question 5	How the dragon felt when the fairy appeared?
Question 6	The end of the story was:

contribution. However, it is insufficient, to assess the overall qualitative improvement to the user experience and narrative intelligibility. This experiment, more concerned with ecological validity, addresses this later situation. Like the previous experiment, a single research question is proposed;

1. What is the overall qualitative impact of music, in the user's experience and narrative intelligibility?

Music is expected to contribute for an overall improvement in the users experience with I-Shadows.

6.3.1 Methodology

Design

This experiment's goal was to qualitatively assess the overall improvement of the user's interaction. Each participant was asked to answer a small questionnaire - appendix H.2 and table 6.7 - about the emotions expressed by the characters. Each question in the questionnaire constituted a dependent variable, in a total of six. At the exception of the second and sixth questions, all other were open questions, i.e., the participants were allowed to freely describe their perceptions. In this experiment one single independent variable was manipulated, the presence or absence of music accompaniment.

Participants

There was a total of 24 Portuguese boys and girls, in the scholar age between 8 and 9 years old. No further demographic factor was measured. All 24 participants attend the same school.

Apparatus

This experiment had a similar set-up to the previous experiment, a single computer connected to a video projector projected the story in a withe wall. There were also two experimenters with the same tasks that were attributed in the previous experiment.

Procedure

This experiment consisted in the projection of a full story, with and without music accompaniment, to two different test groups. The small narrative had four key points that were taken as reference points to take the questions on. The first key point in the story, features a boy and a girl playing and jumping together, the first and second questions refer to this story point. The second important moment is the appearance of a dragon, attacking the happy boy and girl, question 3 refers to this moment. The third key point, corresponds to the story climax, when fairy appears to fight back the dragon, the intelligibility of this moment is measured in questions 4 and 5. Finally, the fourth and last story key point, corresponds to the resolution of the conflict between the fairy and the dragon, in favour of the last one, the perception of the scene overall mood is measured by question 6. After the story's projection the participants were asked to answer a small six-question questionnaire available in appendix H.2. Besides the questionnaire, the experimenter interacting with the children had an informal conversation were each participant was allowed to make their own observations about the story. The participants were divided in groups of six children, in a similar way than the previous experiment. The reasons for such division are also similar, and children were also told to not comment the test procedures to their colleagues, waiting their turn to take the test. The first two groups assisted to a silent story, while the other two watched it with music accompaniment.

6.3.2 Results

The collected data for this experiment consists of free descriptions mainly. Due to this fact, the obtained results are not statistically interpreted, the exception are questions 2 and 6, that were measured in a bipolar five-point scale ranging from very happy to very sad, and which results are summarised in table 6.8.

The first question, "What were the boy and girl doing, at the beginning of the story?", was a simple measure of the general story's intelligibility regardless of the affective states expressed by the characters, i.e., it acted as a control question, serving the purpose of understanding if the children were not having any difficult to interpret the story. In the condition without music accompaniment, 91,67% of the participants said that the boy and girl were playing while 8,33% said they were running, in the condition of music accompaniment 66,67% said they were playing while 25,00% referred that they were jumping. This answer percentages suggest that most participants did not have any trouble to interpret the first moments of the story, and thus their answers to the next questions are expected to be accurate and not affected by a miss understanding of the story mechanisms.

Question 2, "How the boy and girl felt, at the beginning of the story?", was asked against the first story key point, i.e., the initial joyful scene featuring a boy and a girl playing. This question was measured using a five-point bipolar scale, with the extremes of, very happy and very sad. For the condition without music accompaniment 75,00% of the participants rated the scene as being very happy and 16,67% evaluated it as happy, 8,33% were neutral, with music accompaniment the same 75,00% rated the scene as very happy while the remaining 25,00% classified it as happy. This results suggest that in both situations, the participants correctly perceived the overall mood of the scene; however, the usage of music seems to have been slightly clarifying by transforming the 8,33% neutral classifications into happy classifications.

Question 3, "How the boy and girl felt when the dragon appeared?", measured the perceived emotional state of the boy and girl when the dragon attacked by surprise. In the story's projection without music accompaniment 66,67% of the participants said that they were frightened and 33,33% said that they were sad. When the story was projected alongside with accompanying music, the number of participants perceiving the emotional states as frightened increased to 75,00% while 25,00% referred they felt badly. Although these 25,00% can not be accounted to have correctly identified the characters emotions, the percentage of correct identifications increased by 9,23%, suggesting the same clarification effect detected in the previous question.

Question 4, "How the fairy felt when she entered the story?", and question 5, "How the dragon felt when the fairy appeared?", referred to the same key point, the story climax when a fairy appears to fight the dragon and protecte the boy and girl. In the condition without music accompaniment 50,00% of the participants said that the fairy felt happy, and only 25,00% have correctly identified the fairy's emotion as anger, the remaining participants said the fairy was frightened. In what respects to the dragon's emotion 50,0% of the participants correctly identified it as angry, 41,33% said that the dragon was frightened, and 8,33% said that he was feeling happy. The usage of music altered the participants perceptions significantly, 41,67% have correctly identified the fairy's emotion as anger while 25,00% referred she was feeling badly, the remaining participants classified it as frightened, sad and happy. While a significant part did not perceive the fairy's emotion as expected, the 41,67% of participants that did so represents an increase of 15,00% over the number of participants that have also correctly identified the fairy's emotion without music accompaniment. The perception of the dragon's emotions with music accompaniment has also changed with 58,33% of the participants saying that the dragon was angry, a 8,33% improvement over the condition without music accompaniment, while 25,00% and 16,67% said that the dragon was feeling badly and scared, respectively.

Finally, question 6, "The end of the story was?", measured the perception of the story ending overall mood, when the fairy was defeated by the dragon. In the condition without accompaniment music 16,67% of the participants classified the end as very happy, while 16,67% and 66,67% classified it as sad and very sad respectively, as expected. When music was used to accompany the story only 8,33% of the participants still classified the story ending as very happy, on third classified it as neutral and other third as sad, while 25,00% perceived the end as very sad. These results suggest that some

Table 6.8: Frequencies for the scale-based questions, 2 and 6

Questions	Answers	No music		Music	
		Absolut	Percentage	Absolut	Percentage
Question 2	Very happy	9	75,00%	9	75,00%
	Happy	2	16,67%	3	25,00%
	Neutral	1	8,33%	0	0,00%
	Sad	0	0,00%	0	0,00%
	Very sad	0	0,00%	0	0,00%
Question 6	Very happy	2	16,67%	1	8,33%
	Happy	0	0,00%	0	0,00%
	Neutral	0	0,00%	4	33,33%
	Sad	2	16,67%	4	33,33%
	Very sad	8	66,67%	4	25,00%

participants had doubts on classifying the story ending preferring the neutral option, and that a larger group preferred the option of sad instead of very sad, nevertheless, the number of participants classifying it as very happy decreased.

6.3.3 Discussion

This experiment consisted of open descriptive questions, which do not provide suitable data for advanced statistical treatment, but it does provides very ecological descriptions of the user's perceptions, without forgetting the participants young age. The experimenters noted that children where more involved in this experiment than with the previous, probably because they were allowed to freely express their considerations. The analysis of the collected data, suggests that most participants did in fact understood the main guidelines of the story in both experimental conditions, with and without music accompaniment. Nevertheless, the usage of music seems to have contributed to the clarification of some story key points, namely the boy's and girl's emotional state when they were attacked by the dragon, the fairy's anger when she appeared to defend the boy and girl, and the unusual sad ending, with the defeat of the fairy. The results together with the experimenters observations and participant feedback, support the hypothesis that music has a catalytic effect on the improvement of the user's experience in the context of a real story performance. One can question the objectivity of these elements, which are indeed subjective, but these results do not pretend, in anyway, to serve the same purpose of objective statistic measures, they rather preserve ecological validity over data objectivity. It would be very difficult to design a strictly scientific experiment to assess these questions which are inherently complex, thus, this was a conscious option of the experimenters counterbalanced by the previous experiment which at this respect preserves objectivity at the expense of ecological validity.

Chapter 7

Conclusions

"Music begins, where speach ends."

Ernst Hoffmann

Sixty years after Hiller had thrown the rock into the lake with the "Illiac Suite", the question is not, if the computer is able to compose music any longer, but how computer generated music can be used to improve modern computer applications, such as, virtual and interactive environments. In the forefront of these applications, there is the usage of music in affective environments to serve communication, expressive or purely aesthetic purposes. The field of musical emotions has also changed radically in the advent of this century. After decades of neglect, and in spite of a promising start with Hevner's pioneer [32, 33] studies, a renewed interest in musical emotions, emerges in Psychology, Musicology and others disciplines. Influential and important literature such as, the Music and Emotion - Theory and Research book, edited by Juslin and Sloboda [39], laid the ground to the publication of articles and conference communications, sharing the very same motivation synthesised in the following exclamation: "Hey, music can elicit emotions!". How is it able to do it, and what can we do with that?". Born from the intersection of these two fields, and benefiting from favourable winds, the question addressed in this work was the second, "What can we do with music's expressive abilities?".

7.1 Contribution

The I-Sounds project is a practical application of musical emotions to improve the user's interaction experience in affective systems. The extensive literature review, presented in chapter 2 can be regarded as an anthologic perspective over the research on musical emotions, strongly influenced by Juslin and Sloboda [39]. Although, it goes a step ahead and proposes a new perspective, over one of the least studied music structural factors, rhythm. The new perspective is contributed by the hands of a percussionist, with an extensive knowledge of rhythm and its dynamics. The research on musical emotions is above all a multidisciplinary problem and musicians should play an important role on

their study; they have the keys for an extensive empirical knowledge base, about affective composition and affective performance, a precious source waiting to be explored and discovered. The chapter would not be complete without an insight over the field of computer music. That insight is based in the analysis of recent composition and performance systems, developed with a variety of objectives, some very distant from I-Sounds, but enriching for the system itself and useful to provide an overall look on how music is being used to improve the interaction between humans and machines. Finally, it would be unfair to neglect the old and always reinventing art of film music.

One of the objectives of this work, and stated in chapter 1, was the definition of suitable development tools for emotion-based composition algorithms. The proposed architecture in chapter 3, and its respective implementation under the name of I-Sounds, provides such a suitable framework to develop and explore the composition of affective music. This work, felt, in its embryo stage, the lack of supporting tools that would allow to concentrate the efforts on the development of a composition algorithm. It soon become an objective, to develop such tools and provide a concrete implementation that could support this work, but also future researches. Far from being limited to algorithm development, this framework, provides the necessary run-time components to integrate and run those algorithms with affective systems that require the strings of a violin, the keys of a piano, or a pair of drum sticks. Section 4.6 discussed the integration of I-Sounds and I-Shadows. The integration process, validated, I-Sounds as an effective and easy solution to incorporate, music and sounds, with systems which were not initially designed for such purposes. The usage of the I-Sounds framework, in D3S, an ongoing project for an affective music performance system, reflects the I-Sounds flexibility and its utility for other research projects.

Rather than the exclusive product of this project, the I-Sounds framework was used to develop AMADEUS, a composition algorithm based on rhythm and diatonic modes. AMADEUS, addresses the second research question presented in chapter 1. The development of AMADEUS was strongly influenced by Lopes [56] and his theory about the properties of rhythm and metre. Lopes's theory was originally developed with analytical purposes, but AMADEUS applies it to the composition of affective rhythmic sequences. This required the formalisation of the theory's abstractions such as, the mathematical function for the quantification of pulse salience, a fundamental instrument for rhythm composition in AMADEUS. The definition of this formalisations, is a successful example of cooperation between research fields to explore the application of musical emotions in the improvement of interactive systems. The development of AMADEUS, implied the formulation of a set of hypothesis on the relation between the rhythm perceptive qualities of pulse salience and kinesis and the expression of happiness, sadness, anger and fear. While the evaluation discussed in chapter 6 confirms some of these hypothesis, it also reveals the limitations of rhythm and mode, on the expression of anger and fear. Additional research is required on the expression of these two emotions - more on this later.

The present work was also interested in assessing the contribution of the auditive channel for the improvement of user's experience in interactive systems. Pursuing this interest, stated in the third objective referred in chapter 1, AMADEUS, supported in the I-Sounds framework, was integrated

in I-Shadows, an interactive drama application designed for young children. The main goal was to qualitatively and quantitatively measure the impact of music, in what it was, a pure visual interaction.

Two other experiments were presented in chapter 6, the first evaluated the success rate in the recognition of the emotions exhibited by an animated character, while the second measured the improvement in the user's perception and intelligibility of a full narrative performed by I-Shadows. The results emerging from these two experiments are encouraging. The usage of music had greatly improved the recognition rate of happiness and sadness, reaching 100% for the first. Anger and fear obtained mixed results, which is an evidence of the extreme proximity of these two emotions, and their consequent similarity in terms of musical parameters. Nevertheless, the results, especially for anger, could have been improved with another refinement of AMADEUS parameters and with the eventual selection of other segments among the whole set composed by AMADEUS. These marginal results can not be conclusive but a stimulus for further study. The results obtained in the recognition of happiness and sadness are the most important evidence supporting the belief that it is possible to achieve better results for anger and fear. The third experiment had positively assessed the influence of music, in the general improvement of the user's experience with I-Shadows. The ecological data collected in the test and the experimenter's observations support this same idea. The young participants, with an age ranging between 8-9, demonstrated a greater interest and commitment with the test when there was music accompaniment. Although the music is at this stage, very rudimentary from an artistic point of view, the children did not demonstrate any type of discomfort with it.

7.2 Future Work

Rather than a complete solution, this work is above all, a proof of concept. To accomplish this its vision, the one of an affective virtual composer, a continued research activity expanding far beyond from the author's effort in this thesis, is fundamental. I-Sounds and AMADEUS are just an initial step towards the virtual composer with an affective enabled mind. An enormous unknown is available for new explorers. In realistic and rational terms, a full virtual composer able to match the flexibility and creativity of their human counterparts, is likely a work of decades. At this time, and attending to the pre-paradigmatic state of musical emotions, it would be naive to think otherwise.

AMADEUS focused on the properties of rhythm, and while good results were achieved, specially in the expression of happiness and sadness, questions remain open in what respects to anger and fear. Is rhythm insufficient to express such close emotions as this work suggests, or is there an ignored and sublime aspect that can disprove that suggestion? While there is some space left to study rhythm and its usage in AMADEUS, the diatonic modes seem to be fairly understood. However, far from being an exclusive result of rhythm and mode, music has many more factors that can be considered. In what respects to other structural factors there is no good reason to pick one, or another, instead of this or that one. This decision should be made by the researcher whiling to continue the work. Nevertheless, some can be suggested, the usage of other tonalities and modes, a wider set of pitches and the

manipulation of tempo and dynamics are promising options. Irrespective from the future researcher's choice, a bigger concern and care should be put in the control of multiple factor interactions, e.g., if the researcher decides to manipulate the parameter of tempo, the considerations and results about the rhythm property of kinesis developed in this work are likely to be affected.

The I-Sounds framework proved to be a good solution to support the development of affective composition algorithms and their integration with other systems. While some architectural improvements and functional extension might be possible and desirable, the actual revision is likely to serve most research needs. Of course, this does not affects the possible development of advanced emotional and composition models as internal representations for I-Sounds, but that should not affect the system's generality and extensibility. These features should always be regarded as concrete implementations that can fit, or not, the researcher needs, and therefore should be optional. The system's core can also be improved, namely its computational optimisation, which is likely to be mandatory, as the composition algorithms grow in complexity. Nevertheless, the main idea and goal, is that researchers wishing to concentrate on musical emotions and the development of affective composition algorithms, should not be required to expend more than a third of their available time, in the development of support logic. The system's optimisation is likely to request for special attention and an independent work focused on that same task.

7.3 Final considerations

The door had just been open. The study of musical emotions is expected to enlighten yet undiscovered aspects of our emotions, contributing to a better understanding of such important aspect that defines us as humans. Consequently, we will better understand the music itself, a manifestation of the human intellect that although ubiquitous since the pre-historic ages, remains an incredible pot of surprises. The research context, both in Psychology and Musicology, seems to be the most favourable since the appearance of both sciences. The multi-disciplinary collaboration is fortunately a trend, new and different insights over the questions will only help but to better solve them. Of the same paramount importance, is to answer the question of how this new knowledge can be used to improve other areas. Undoubtedly, the emerge of new interactive systems requires for multi-modal interaction, there is no technological reason to be differently. Of great relevance is the user's awareness of these possibilities, and his demand for its usage in the interactive systems of today, such as ADA, referred in chapter 2. The scope of this discussion, surely expands farther than this work's objectives; however, this thesis was thought to be a step towards the future of musical emotions and their practical usage to achieve better interactive experiences, but many more steps are needed, and the field claims for more highly motivated researchers from all disciplines, to join the efforts towards the virtual composer.

Appendix A

Music Terminology Glossary

articulation The performance technique affecting the continuity of a single note or the transition between multiple notes. Articulation types or marks, include staccato, staccatissimo, accent, sforzando, and legato, among others. See the glossary entries for staccato and legato. 14, 15, 26

ascending melody A sequence of notes in ascending order of pitch. 13

atonality The absence of tonality, i.e., the lack of a tonal centre in the pitch hierarchy. Atonal music emerged in the beginning of the 20th century as opposed to the tonal rules used from the 17th until the 19th century in western music. 14

bar A synonym for measure, see the respective entry. 24

beat The basic metric unit. Each tick of a metronome denotes a beat. 14, 17, 18

BPM beats per minute. 11

chord A group of notes played simultaneously. 12–14, 21

chromatic tonality A type of tonality in which the set of diatonic pitches is expanded to include five additional chromatic pitches. The result is a set of twelve pitches equally separated by semitones, that can actually be chosen as tonal centres. 14

clef A notation symbol that gives the name and pitch to each of the staff's notes, placed in its spaces and lines. 30

crescendo The term crescendo is used in music notation to express an increase of sound loudness along a time period. The opposite of diminuendo. 13

descending melody A sequence of notes in descending order of pitch. 14

diatonic scale Seven-note scales comprising five whole-tones and two half-tones, laid in a manner that maximizes the distance between the two half-tones. 12

diminuendo The term diminuendo is used in music notation to express a decrease of sound loudness along a time period. The opposite of crescendo. 13

dynamics The softness or loudness of musical notes. The term is also applied to other stylistic aspects of music performance such as, legato and staccato. The functional property of velocity is also part of music dynamics. 10, 12–15, 26

eighth note A note with an eighth of a whole note's duration, represented by filled-in ovals with a single flag straight stem. In classical terminology eighth notes are named quavers. 18

half tone A pitch difference between two notes consisting of exactly an half of a whole tone. Its exactly meaning in frequency is dependent on the temper system. Half-tones are sometimes named Semitones. 30

harmonic formula One of the formulas of the minor diatonic mode, the other two being, the natural and melodic formulas. Each formula sounds distinctly. 12

harmonic interval The pitch interval between two music notes simultaneously played. 14

harmony Harmony comprises the use of different pitches simultaneously, from which chords are an example. Harmony is usually referred as music's vertical dimension. 13, 14, 17, 20

legato A stylistic property of music performance. The transition between notes played legato should sound smooth, i.e., no silence between each note. The opposite of staccato. 15

loudness Loudness is the quality of sound related with its physical strength, amplitude. Loudness may also be affected by other sound properties. 12, 13, 26

measure A group of beats usually, 2, 3 or 4. Measures organise the piece's beats into logical segments with equal duration. The number of beats in each measure, rarely varies inside the same piece. The term bar, is a synonym, frequent in American English. 14

melodic interval The pitch interval between two successive notes. 13, 14

melody A sequence of notes building musical phrases. Although the term applies to any sequence of pitches, often it is applied to refer those sequences that obey to certain pitch succession rules, being perceived as consonant and pleasing. 13, 14, 16, 17, 20, 29

meter The measurement of a musical line into measures of stressed and unstressed beats. Rhythm concerns patterns of duration while metre involves the perception of natural divisions as abstracted from rhythm. 14–18

minor second An half-tone interval between two notes. 13

modality The usage and conception of musical modes. Modality is usually compared to tonality and atonality, as an alternative music system. 14

mode The interval pattern between each of the scale's notes. Modes entail a distinctive physiognomy on scales, eliciting its character or mood. 12, 14

mood mark A set of notation terms used to express the character of a piece. These are usually written in Italian and include words such as; agitato, dolce, maestoso, vivace, etc. 12

musical form The type of composition, e.g., symphony, or the structure of a particular piece, e.g., sonata. The term can also refer to the usage of fine-grained elements of music structure, such as; repetition, condensation, etc. 30

narrow melody A sequence of notes with low amplitude between the lowest and higher pitches. 30

note The sounding element of music. Its two fundamental properties are pitch and duration, rising the melodic and rhythmic music dimensions, respectively. 11–13, 15, 16, 18, 24

octave The interval between one note, and other with half or double of its frequency. Comprises five whole-tones and two half-tones. 13

pitch Pitch refers to the fundamental frequency of a sound. Each of the seven musical notes have different frequencies and therefore different pitches. Pitch is the measure used in scales to order sounds from low to high. 13, 16, 17, 26

quarter note A note with a quarter of a whole note's duration, represented by filled-in ovals with straight flagless stems. In classical terminology, quarter notes are named, crotchets. 11, 18

rest An interval of silence in music. Such as notes, rests are structured in a fractional hierarchy, e.g., the quarter rest has a quarter of the duration of the whole rest. 11, 15, 16

rhythm The variation of the length and accentuation of a series of notes and rests. 14, 16–18, 22, 29

scale A set of musical notes, ordered in pitch providing a measure of musical distance between notes. Scales are to the composer, which colour palette's are to the painter. 12

score The written - hand-written or printed - notation of a music piece. 10–13, 29

staccato A stylistic property of music performance. Notes played staccato should sound in a detached manner, i.e., a very short silence fills the last fraction of the note's duration. The opposite of legato. 15

staff A set of five lines and four spaces to represent notes according to their pitch. 30

tempo The pace of a piece sometimes expressed in beats per minute. From the Latin *Tempus*. 10–12, 14, 18, 20, 26

tempo mark A set of notation terms used to qualitatively indicate the piece's tempo. These terms are usually written in Italian but other languages may also appear. The set includes terms such as; presto, allegro, moderaro, andante, adagio, largo, lento, etc. 30

timbre The quality of sound allowing the distinction among different instruments. Timbre results from the spectrum and envelope physical characteristics. Sound quality or sound colour are synonyms of timbre, in psycho-acoustics. 30

time signature A notation mechanism to define how many beats exist in a measure and the beat value. It is usually expressed through a fraction without the horizontal slash between the numbers. The numerator defines the number of beats and the denominator the beat value. 16, 20

tonality A music system in which pitch relationships are established around a tonal centre or tonic. Tonality is expressed through the term key, e.g., this piece is in the key of G major, meaning the piece is arranged around the tonic of G. 14

tone harmonics An harmonic is a component frequency of a signal, which is an integer multiple of its fundamental frequency. The sounds produced by an instrument can have multiple harmonics. Musicians usually use the term overtone to refer to each of those harmonics. 15

tonic The first note of a musical scale around which the pitch hierarchy is organised. 30

whole tone A pitch difference between two notes, consisting of two semitones. Its exactly meaning in frequency is dependent on the temper system. 30

wide melody A sequence of notes with high amplitude between the lowest and higher pitches. 30

Appendix B

Detailed architecture

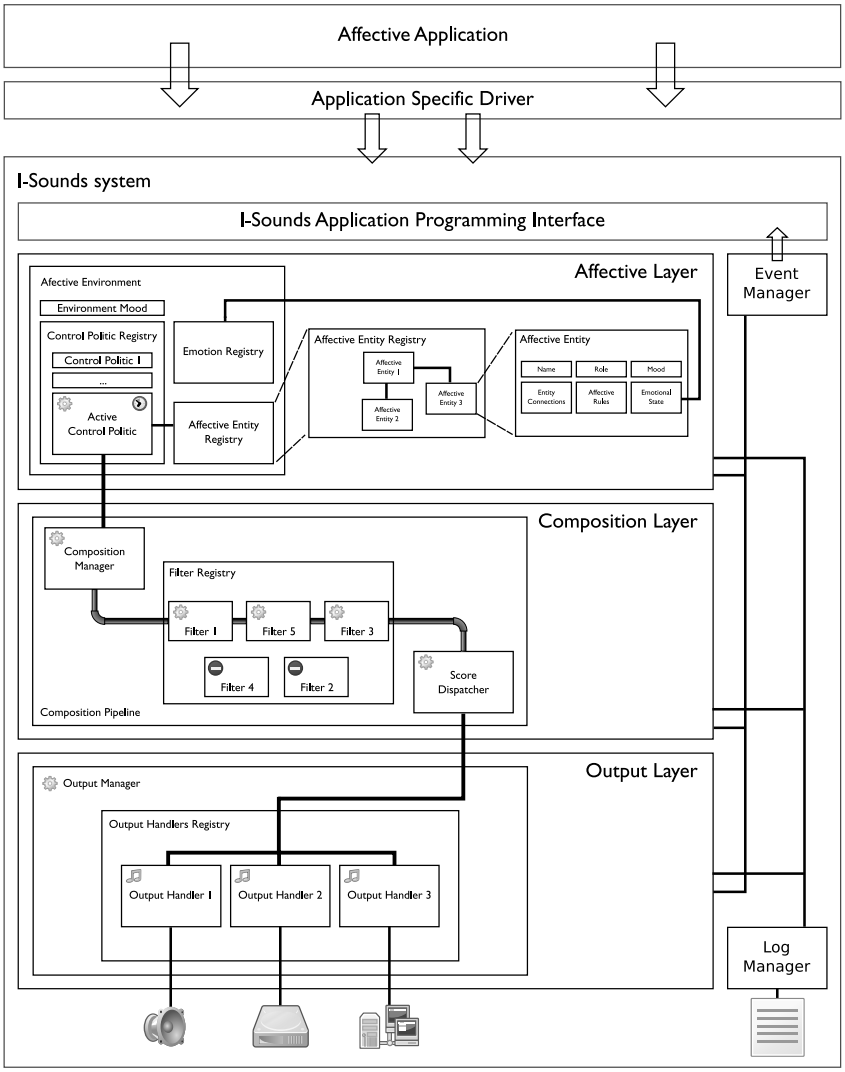


Figure B.1: Detailed block diagram of the proposed architecture

Appendix C

I-Shadows/I-Sounds XML Protocol

Message 1 details			
No.	Type	Code	Description
1	control	1	Register new affective entities

```
<message type="control"code="1">
        <content>
                <actor name="Bert"role="the␣banana"/>
                <actor name="Ernie"role="the␣orange"/>
        <content>
</message>
```

Message 2 details			
No.	Type	Code	Description
2	control	2	Remove affective entities

```
<message type="control"code="2">
        <content>
                <actor name="Bert"/>
                <actor name="Ernie"/>
        <content>
</message>
```

Message 3 details			
No.	Type	Code	Description
3	control	3	Establish a connection between affective entities

```
<message type="control"code="3">
 <content>
  <actorRelation source="Bert"destiny="Ernie"description="roomy"value="10"/>
  <actorRelation source="Ernie"destiny="Bert"description="roomy"value="10"/>
 <content>
</message>
```

Message 4 details			
No.	Type	Code	Description
4	info	1	Update the mood value of affective entities

```
<message type="control"code="2">
        <content>
                <actor name="Bert"mood="5"/>
                <actor name="Ernie"mood="5"/>
        <content>
</message>
```

Message 5 details			
No.	Type	Code	Description
5	info	2	Update the emotional state of affective entities

```
<message type="control"code="2">
        <content>
                <actor name="Bert"emotion="joy"value="10"/>
                <actor name="Ernie"emotion="joy"value="10"/>
        <content>
</message>
```

Message 6 details			
No.	Type	Code	Description
6	info	3	Update a connection between two affective entities

```
<message type="info"code="3">
 <content>
  <actorRelation source="Bert"destiny="Ernie"description="roomy"value="10"/>
  <actorRelation source="Ernie"destiny="Bert"description="roomy"value="10"/>
 <content>
</message>
```

Message 7 details			
No.	Type	Code	Description
7	info	4	Update the environment's overall mood

```
<message type="control"code="2">
        <content>
                <environment mood="10"/>
        <content>
</message>
```

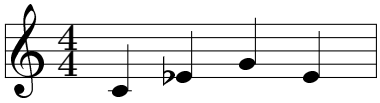
Message 8 details			
No.	Type	Code	Description
8	info	5	Inform about the most active affective entity

```
<message type="info"code="5">
        <content>
                <actor name="Ernie"role="the␣orange"/>
        <content>
</message>
```

Appendix D

Music XML Score Notation Example

A score in traditional notation and its Music XML representation;



Score D.1: AMADEUS-composed score eliciting sadness

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE score-partwise PUBLIC "-//Recordare//DTD MusicXML 2.0 Partwise//EN" "http://www.musicxml.org/dtds/partwise.dtd">
<score-partwise version="2.0">
    <work>
        <work-title>Sequence</work-title>
    </work>
    <identification>
        <creator type="compose">AMADEUS</creator>
    </identification>
    <part-list>
        <score-part id="P1">
            <part-name>Part1</part-name>
        </score-part>
    </part-list>
    <part id="P1">
        <measure number="1">
            <sound tempo="80"/>
            <attributes>
                <divisions>4</divisions>
                <key>
                    <fifths>0</fifths>
                    <mode>major</mode>
                </key>
                <time>
```

```
                    <beats>4</beats>
                    <beat-type>4</beat-type>
                </time>
                <clef>
                    <sign>G</sign>
                    <line>2</line>
                </clef>
            </attributes>
            <note>
                <pitch>
                    <step>C</step>
                    <octave>4</octave>
                </pitch>
                <duration>4</duration>
                <type>quarter</type>
            </note>
            <note>
                <pitch>
                    <step>C</step>
                    <octave>4</octave>
                </pitch>
                <duration>4</duration>
                <type>quarter</type>
            </note>
            <note>
                <pitch>
                    <step>C</step>
                    <octave>4</octave>
                </pitch>
                <duration>4</duration>
                <type>quarter</type>
            </note>
            <note>
                <pitch>
                    <step>C</step>
                    <octave>4</octave>
                </pitch>
                <duration>4</duration>
                <type>quarter</type>
            </note>
        </measure>
    </part>
</score-partwise>
```

Appendix E

I-Sounds Configuration Files

E.1 Core Configuration

```
#################################################################
#        The I-Sounds configuration file
#################################################################

#################################################################
#        I-Sounds bundled jar location
#################################################################
# This path can be relative to the JVM execution directory
# or an absolute reference
I-SoundsJarFile = ./lib/I-Sounds.jar

#################################################################
#        Affective Layer configuration
#################################################################
# These properties configure the AffectiveLayer of I-Sounds
# You can define the default control politic and emotions
# to load when starting the system
AffectiveLayer\:DefaultControlPoliticName = EmotionAverageControl
AffectiveLayer\:DefaultControlPoliticClass = iSounds.affectiveLayer.
    controlPolitics.EmotionAverageControl
AffectiveLayer\:Emotions = iSounds.affectiveLayer.emotions.Joy \
                           iSounds.affectiveLayer.emotions.Distress \
                           iSounds.affectiveLayer.emotions.Anger \
                           iSounds.affectiveLayer.emotions.Fear
```

```
##############################################################
#        Composition Layer configuration
##############################################################
# These properties set the internal buffer capacity of some
# compsing layer elements. This properties can influence
# system behaviour and/or performance
CompositionPipeline\:DefaultComposerName = DummyComposer
CompositionPipeline\:DefaultComposerClass = iSounds.compositionLayer.composer.
    DummyComposer
CompositionManager\:BufferCapacity = 10
ScoreDispatcher\:BufferCapacity = 10

##############################################################
#        Output Layer configuration
##############################################################
# You can configure the default output handler that is
# loaded when the system starts
OutputLayer\:DefaultOutputHandlerName = MidiOutputHandler
OutputLayer\:DefaultOutputHandlerClass = iSounds.outputLayer.outputHandlers.
    MidiOutputHandler
OutputManager\:BufferCapacity = 10

##############################################################
#        MIDI Output Handler resource configuration
##############################################################
# You can configure the MIDI devices that I-Sounds should
# use by default. This setting can also be set via the
# I-Sounds API
MIDIOutputHandler\:MIDISynthesizer = Java Sound Synthesizer
MIDIOutputHandler\:MIDISequencer = Real Time Sequencer

##############################################################
#        LogManager and log facilities configuration
##############################################################
# You can configure the logging system using a configuration file
LogManager\:ConfigFile = ./etc/logging.properties
```

E.2 Logging Facilities Configuration

```
############################################################
#       Default Logging Configuration File
#
# You can use a different file by specifying a filename
# with the java.util.logging.config.file system property.
# For example java -Djava.util.logging.config.file=myfile
############################################################

############################################################
#       Global properties
############################################################
# "handlers" specifies a comma separated list of log Handler
# classes.  These handlers will be installed during VM startup.
# Note that these classes must be on the system classpath.
# By default we only configure a ConsoleHandler, which will only
# show messages at the INFO and above levels.
handlers=

# To also add the FileHandler, use the following line instead.
#handlers= java.util.logging.FileHandler, java.util.logging.ConsoleHandler

# Default global logging level.
# This specifies which kinds of events are logged across
# all loggers.  For any given facility this global level
# can be overriden by a facility specific level
# Note that the ConsoleHandler also has a separate level
# setting to limit messages printed to the console.
.level= INFO

############################################################
# Handler specific properties.
# Describes specific configuration info for Handlers.
############################################################
# default file output is in user's home directory.
java.util.logging.FileHandler.pattern = %h/java%u.log
java.util.logging.FileHandler.limit = 50000
java.util.logging.FileHandler.count = 1
java.util.logging.FileHandler.formatter = java.util.logging.XMLFormatter
```

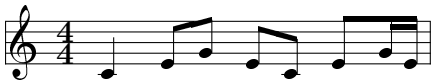
```
# Limit the message that are printed on the console to INFO and above.
java.util.logging.ConsoleHandler.level = INFO
java.util.logging.ConsoleHandler.formatter = java.util.logging.SimpleFormatter


############################################################
# Facility specific properties.
# Provides extra control for each logger.
############################################################
# For example, set the com.xyz.foo logger to only log SEVERE
# messages:
com.xyz.foo.level = SEVERE
```

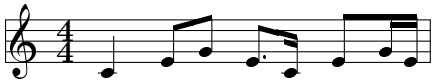
Appendix F

Evaluation Music Scores

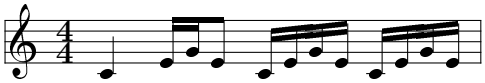
F.1 Happy Segments



Score F.1: Happy segment 1

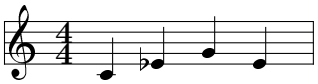


Score F.2: Happy segment 2

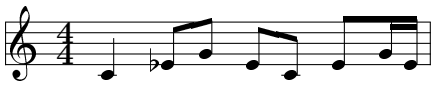


Score F.3: Happy segment 3

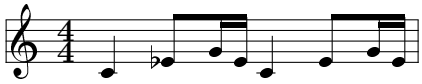
F.2 Sad Segments



Score F.4: Sad segment 1

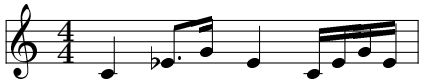


Score F.5: Sad segment 2

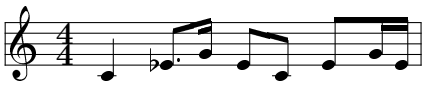


Score F.6: Sad segment 3

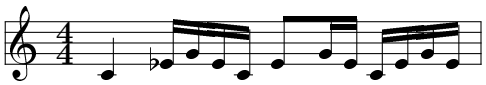
F.3 Angry Segments



Score F.7: Angry segment 1

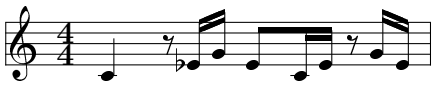


Score F.8: Angry segment 2

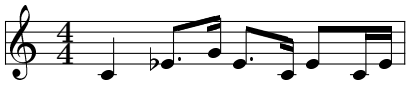


Score F.9: Angry segment 3

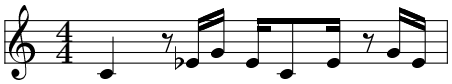
F.4 Frightning Segments



Score F.10: Frightning segment 1



Score F.11: Frightning segment 2



Score F.12: Frightning segment 3

Appendix G

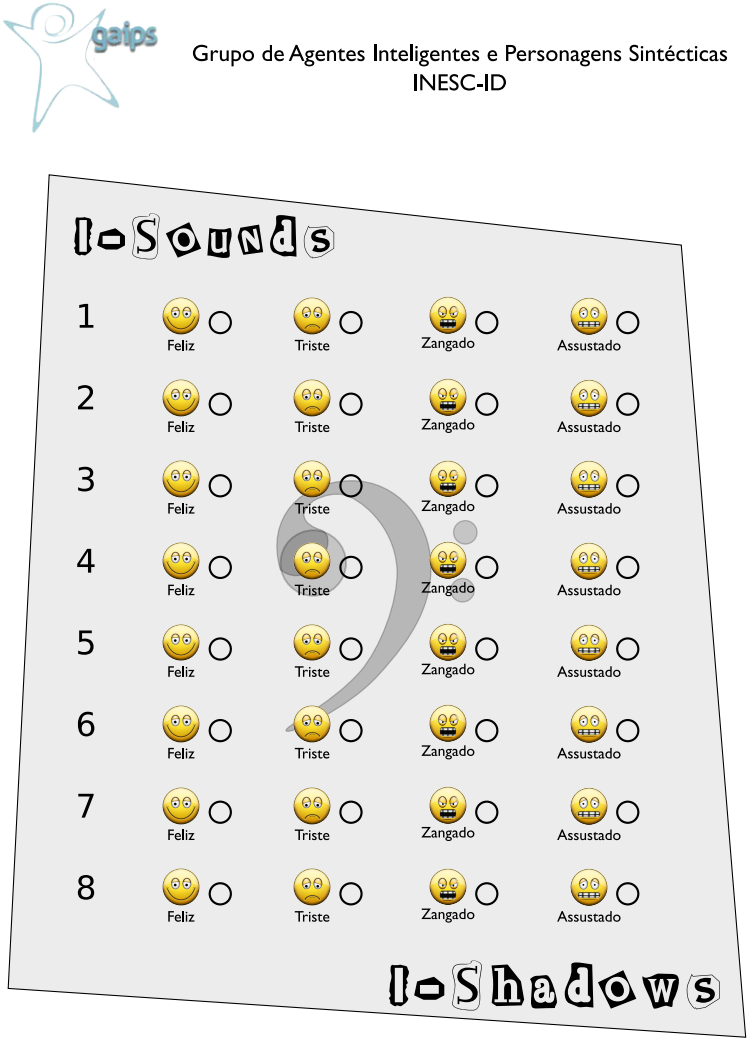
Proposed Salience Intervals

Table G.1: Hypothesized salience intervals for rhythm composition

Emotion		**First beat**				**Second beat**			
	m.p.	1	2	3	4	5	6	7	8
Happiness	min.	5000	0	0	0	3750	1750	2750	750
	max.	5000	0	0	0	4000	1750	3000	750
Sadness	min.	5000	0	0	0	4000	0	2750	750
	max.	5000	0	0	0	4500	0	3000	750
Anger	min.	5000	0	0	0	3750	1750	2750	750
	max.	5000	0	0	0	4250	1750	2750	750
Fear	min.	5000	0	0	0	0	2250	2750	750
	max.	5000	0	0	0	4250	2250	3000	750
Emotion		**Third beat**				**Fourth beat**			
	m.p.	9	10	11	12	13	14	15	16
Happiness	min.	4000	2000	3000	1000	3500	1500	2500	500
	max.	5250	2000	3250	1000	4750	1500	2500	500
Sadness	min.	4250	0	3250	0	3750	0	2500	500
	max.	7250	0	3250	0	5250	0	2750	500
Anger	min.	4750	0	3000	1000	3500	1500	2500	500
	max.	8750	0	3250	1000	4750	1500	2500	500
Fear	min.	0	2250	3000	1000	0	0	2500	500
	max.	5250	2500	3250	1000	3750	0	2500	500

Appendix H

Experimental Evaluation Sheets

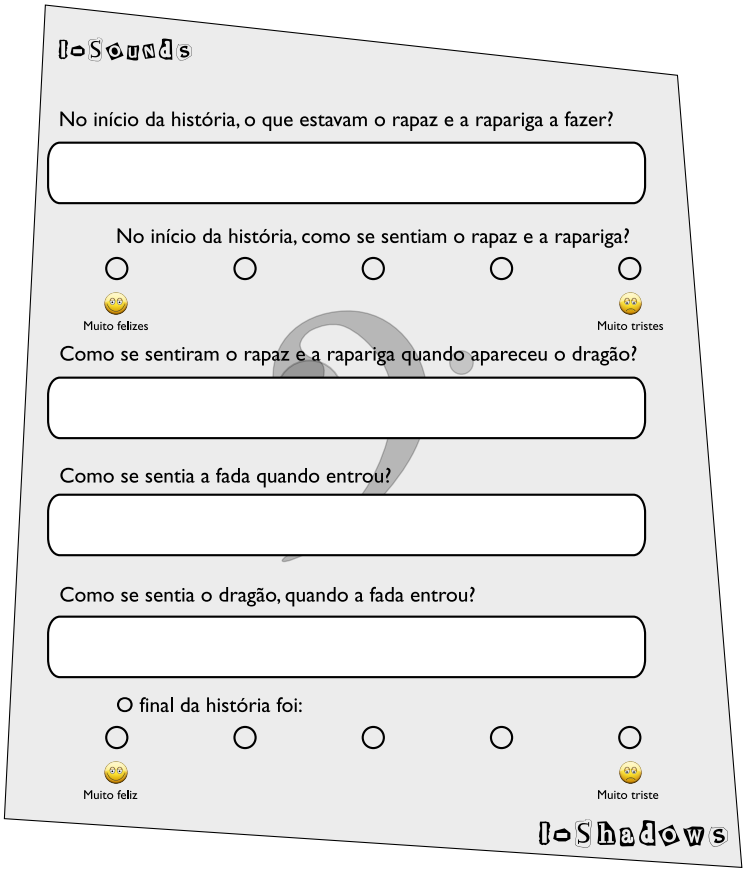

Grupo de Agentes Inteligentes e Personagens Sintécticas
INESC-ID

IoSounds

1	Feliz	Triste	Zangado	Assustado
2	Feliz	Triste	Zangado	Assustado
3	Feliz	Triste	Zangado	Assustado
4	Feliz	Triste	Zangado	Assustado
5	Feliz	Triste	Zangado	Assustado
6	Feliz	Triste	Zangado	Assustado
7	Feliz	Triste	Zangado	Assustado
8	Feliz	Triste	Zangado	Assustado

IoShadows

Figure H.1: The report sheet for the emotion recognition experiment


IoSounds

No início da história, o que estavam o rapaz e a rapariga a fazer?

No início da história, como se sentiam o rapaz e a rapariga?

Muito felizes

Muito tristes

Como se sentiram o rapaz e a rapariga quando apareceu o dragão?

Como se sentia a fada quando entrou?

Como se sentia o dragão, quando a fada entrou?

O final da história foi:

Muito feliz

Muito triste

IoShadows

Figure H.2: The report sheet for the story projection experiment

Bibliography

[1] Balkwill, L. L., and Thompson, W. F. A cross cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music perception*, 17 (1999), 43–64.

[2] Becker, J. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, ch. Anthropological Perspectives on Music and Emotion, pp. 135–160.

[3] Behne, K. E. *Der Einfluss des Tempos auf die Beurteiling von Musik.* Arno Volk Verlag, Köln, Germany, 1972.

[4] Brisson, A., Fernandes, M., and Paiva, A. Children as affective designers. In *HUMAINE WP9 workshop* (Kista, Sweden, 2006).

[5] Chung, J., and Vercoe, G. S. The affective remixer: personalized music arranging. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems* (New York, NY, USA, 2006), ACM, pp. 393–398.

[6] Cohen, A. J. Introduction to the special volume on the psychology of film music. *Psychomusicology*, 13 (1994), 2–8.

[7] Cohen, A. J. *Music mind, and science.* Seoul National University Press, Seoul, South Korea, 1999, ch. The functions of music in multimedia, pp. 53–69.

[8] Cohen, A. J. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, ch. Music as a source of emotion in film, pp. 249–272.

[9] Cook, N., and Dibben, N. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, ch. Musicological Approaches to Emotion, pp. 45–70.

[10] Cope, D. *Computers and Musical Style.* A-R Editions, Madison, MI, 1991, 1991.

[11] Cope, D. *Music: Computer Synthesis of Musical Style.* MIT Press, Cambridge, MA, USA, 2001.

[12] Copland, A. *Our new music.* McGraw-Hill, New York, 1941.

[13] Crowder, R. G. Perception of the major/minor distinction: Iii.hedonic, musical, and affective descriminations. *Psychomusicology*, 23 (1985), 314–316.

[14] Davies, J. B. *The psychology of music.* Hutchinson, London, 1978.

[15] Davies, S. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, ch. Philosophical Prespectives on Music Expressiveness, pp. 23–44.

[16] Denora, T. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, ch. Aesthetic Agency and Musical Practice: New Directions in The Sociology of Music and Emotion, pp. 161–180.

[17] Deutsch, D., Ed. *The psychology of music*, 2nd edition ed. Academic Press, San Diego, CA, 1999.

[18] Dowling, W., and Harwood, D. L. *Music cognition.* Academic Press, New York, 1986.

[19] Drake, C., Dowling, J., and Palmer, C. Accent structures in the reproduction of simple tunes by children and adult pianists. *Music Perception*, 8 (1991), 315–334.

[20] Drake, C., and Palmer, C. Accent structures in music performance. *Music Perception*, 10 (1993), 343–378.

[21] Dutta, S., and Kanungo, R. N. *Affect and Memory: A reformulation.* Pergamon Press, New York, 1975.

[22] Eladhari, M., Nieuwdorp, R., and Fridenfalk, M. The soundtrack of your mind: mind music - adaptive audio for game characters. In *ACE '06: Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology* (New York, NY, USA, 2006), ACM, p. 54.

[23] Farbood, M., and Schoner, B. Analysis and synthesis of palestrina-style counterpoint using markov chains. In *International Computer Music Conference* (2001).

[24] G., P., and Wiggins, G. A genetic algorithm for the generation of jazz melodies. In *Procedings of the STeP'98, Jyväskylä, Finland, 1998* (1999).

[25] Gabrielsson, A. Adjective ratings and dimension analysis of auditory rhythm patterns. *Scandinavian -hournal of Psychology* (1973), 244–260.

[26] Gabrielsson, A. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, ch. Emotions in Strong Experiences with Music, pp. 431–449.

[27] Gabrielsson, A., and Juslin, P. N. Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, 24 (1996), 68–91.

[28] Gabrielsson, A., and Lindström, E. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, ch. The Influence of Musical Structure on Emotional Expression, pp. 223–248.

[29] Gerardi, G. M., and Gerken, L. The development of affective responses to modality and melodic contour. *Music Perception* (1995), 279–290.

[30] Gotlieb, H., and Konecni, V. J. The effects of instrumentation, playing style, and structure in the goldberg variations by johann sebastian bach. *Music Perception*, 3 (1985), 87–102.

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

[32] Hevner, K. Experimental studies of the elements of expression in music. *American Journal of Psychology*, 48 (1936), 246–268.

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

[34] Holland, J. H. *Adaptation in Natural And Artificial Systems.* University of Michigan Press, Ann Arbor, MI, US, 1975.

[35] Imberty, M. *Entendre la Music.* Dunod, Paris, 1979.

[36] J., M., and Verschure, P. Roboser: A real-world composition system. *Computer Music Journal 29*, 3 (2005), 55–74.

[37] Jewell, M. O., Nixon, M. S., and Prügel-Bennett, A. Cbs: A concept-based sequencer for soundtrack composition. In *Proceedings of the Third International Conference WEB Delivering of Music (WEDELMUSIC' 03)* (2003), IEEE.

[38] Juslin, P. N. Perceived emotional expression in synthesized performances of a short melody: Cpaturing the listener's judgment policy. *Music Scientiae*, 1 (225-256).

[39] Juslin, P. N., and Sloboda, J. A., Eds. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001.

[40] Juslin, P. N., and Sloboda, J. A. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, ch. Psychological Perspectives on Music and Emotion, pp. 71–104.

[41] Ka-Hing, J. M., Sze-Tsan, C. C., Kwok-Fung, C. S., and Chi-Fai, H. M. Emotion-driven automatic music arrangement. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Research posters* (New York, NY, USA, 2006), ACM, p. 108.

[42] Karno, M., and Konecni, V. J. The effects of structural interventions in the first movement of mozart's symphony in g minor k. 550 on aesthetic preference. *Music Perception*, 10 (1992), 63–72.

[43] Kim, S., and André, E. A generate and sense approach to automated music composition. In *IUI '04: Proceedings of the 9th international conference on Intelligent user interfaces* (New York, NY, USA, 2004), ACM, pp. 268–270.

[44] Kivy, P. *Sound Sentiment: An Essay on the Musical Emotions, Including the Complete Text of The Gorded Shell.* Temple University Press, 1989.

[45] Kleinen, G. *Experimentelle Studien zum musikalischen Ausdruck.* Universität Hamburg, Hamburg, Germany, 1968.

[46] Konecni, V. J. *Cognitive processes in the perception of art.* North-Holand, Amsterdam, 1984, ch. Elusive effects of artists "messages", pp. 71–93.

[47] Konecni, V. J., and Karno, M. Empirical investigations of the hedonic and emotional effects of musical structure. *Musikpsychologie, Jahrbuch der Seutschen Gesellschaft für Musikpsychologie*, 11 (1999), 119–137.

[48] Kratus, J. A developmental study of children's interpretation of emotion in music. *Psychology of Music*, 21 (1993), 19.

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

[50] L., K. C. An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*, 51 (1997), 336–353.

[51] Le Groux, S., Manzolli, J., and Verschure, P. Vr-roboser: real-time adaptive sonification of virtual environments based on avatar behavior. In *NIME '07: Proceedings of the 7th international conference on New interfaces for musical expression* (New York, NY, USA, 2007), ACM, pp. 371–374.

[52] Legaspi, R., Hashimoto, Y., Moriyama, K., Kurihara, S., and Numao, M. Music compositional intelligence with an affective flavor. In *IUI '07: Proceedings of the 12th international conference on Intelligent user interfaces* (New York, NY, USA, 2007), ACM, pp. 216–224.

[53] Lerdahl, F. Cognitive constraints on compositional systems. *Contemporary Music Review 6* (1992).

[54] Lewis, M., and Haviland-Jones, J. M., Eds. *Handbook of emotions*, 2nd edition ed. The Guilford Press, New York, 2000.

[55] Lindström, E. Impact of melodic structure on emotional expression. In *Proceedings of the third triennial ESCOM Conference, Uppsala, June 1997* (Uppsala, Sweden, 1997), Uppsala University, pp. 292–297.

[56] Lopes, E. *Just In Time: Towards a theory of rhythm and metre.* PhD thesis, University of Southampton, UK, 2003.

[57] Maher, T. F., and Berlyne, D. E. Verbal and exploratory reponses to melodic musical intervals. *Psychology of Music*, 10 (1982), 11–27.

[58] Marks, M. *Music and the silent film.* Oxford University Press, New York, 1997.

[59] McCormack, J. Grammar based music composition. In *Complex Systems* (1996).

[60] Meyer, L. B. *Emotion and Meaning in Music.* University of Chicago Press, London, 1956.

[61] Microsoft Corporation. *Composing music for interactive titles: An overview of DirectMusic Producer*, 1998.

[62] Motte-Haber, H. d. l. *Ein Beitrag zur Klassifikation musikalischer Rhythmen.* Arno Volk Verlag, Köln, Germany, 1968.

[63] Nelson, G. L. Fractal mountains. *Computer Music Currents 10* (1992).

[64] Nielsen, F. V. *Oplevelse av musikalsk spoending [Experience of musical tension].* Akademisk Forlag, Copenhagen, Denmark, 1983.

[65] Numao, M., Takagi, S., and Nakamura, K. Constructive adaptive user interfaces - composing music based on human feelings. In *Proceedings of the AAAI 2002* (2002), AAAI Press.

[66] Oppenheim, Y. The functions of film music. Published in the online magazine Film Score Monthly, http://www.filmscoremonthly.com/features/functions.asp, (last accessed on January, 23rd, 2007).

[67] Ortony, A., Clore, G., and Collins, A. *The Cognitive Structure of Emotions.* Cambridge University Press, Cambridge, MA, USA, 1998.

[68] Palmer, C. *The composer in Hollywood.* Marion Boyars, New York, 1990.

[69] Peretz, I. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, ch. Listen to the Brain: A Biological Perspective on Musical Emotions, pp. 105–134.

[70] Peretz, I., Gagnon, L., and Bouchard, B. Music and emotion: perceptual determinants immediacy, and isolation after brain damage. *Cognition*, 68 (1998), 111–141.

[71] Peretz, I., and Morais, J. Music and modularity. *Contemporary Music Review*, 4 (1989), 279–293.

[72] Prendergast, R. M. *Film music: A neglected art*, 2nd edition ed. Norton, New York, 1991.

[73] Rigg, M. G. What features of a musical phase have emotional suggestiveness? *Publications of the Social Science Research Council of the Oklahoma Agricultural and Mechanical College*, 1 (1939).

[74] Rigg, M. G. The effect of register and tonality upon musical mood. *Journal of Musicology*, 2 (1940), 49–61.

[75] Rigg, M. G. Speed as a determiner of musical mood. *Journal of Experimental Psychology*, 27 (1940), 566–571.

[76] Roads, C. *The Computer Music Tutorial.* MIT Press, Cambridge, MA, 1995.

[77] Rosar, W. Film music and heinz werner's theory of physionomic perception. *Psychomusicology*, 13 (1994), 154–165.

[78] Rowe, R. *Interactive music systems: machine listening and composing.* MIT Press, Cambridge, MA, USA, 1992.

[79] Russel, J. A. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39 (1980), 1161–1178.

[80] Russel, S., and Norvig, P. *Artificial Intelligence: A Modern Approach*, 2nd edition ed. Prentice Hall, 2002.

[81] S., N., and Cesarec, Z. Emotional experience of music as a function of musical structure. *Psychology of Music*, 10 (1982), 7–17.

[82] Scherer, K. R., and Oshinsky, J. S. Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, 1 (1977), 331–346.

[83] Scherer, K. R., and Zentner, M. R. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, ch. Emotional Effects of Music Production Rules, pp. 361–392.

[84] Schottstaedt, W. *Automatic Counterpoint.* MIT Press, 1989.

[85] Schubert, E. *Measurement and Time Series Analysis of Emotion in Music.* PhD thesis, University of New South Wales, Sydney, Australia, 1999.

[86] Schubert, E. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, pp. 393–414.

[87] Sloboda, J. A., and O'Neill, S. A. *Music and Emotion - Theory and Research.* Series in Affective Science. Oxford University Press, New York, 2001, ch. Emotions in Everyday Listening to Music, pp. 415–429.

[88] Smith, L. D., and Williams, R. N. Children's artistic responses to musical intervals. *Journal of Psychology*, 112 (1999), 383–410.

[89] Thayer, J. F. Multiple indicators of affective responses to music. *Dissertation Abstracts International*, 47 (1986), 12.

[90] Thompson, W. F., and Robitaille, B. Can composers express emotions through music? *Empirical Studies of the Arts*, 10 (1992), 79–89.

[91] Tillman, B., and Bigand, E. Does formal musical structure affect perception of musical expressiveness? *Psychology of Music*, 24 (1996), 1–17.

[92] Todd, N. Motion in music: a neurobiological prespective. *Music Perception*, 17 (1999), 115–129.

[93] Unehara, M., and Onisawa, T. Construction of music composition system with interactive genetic algorithm. In *Proceedings of the Asian Design International Conference* (2003).

[94] Wanderley, W. W., Schell, N., and Rovan, J. B. Escher - modeling and performing composed instruments in realtime. In *Proceedings of the IEEE International Conference on Systems Man and Cybernetics, San Diego - CA, USA* (1998).

[95] Wassermann, K. C., and et al. Roboser: An autonomous interactive composition system. In *Proceedings of the International Computer Music Conference* (2000), International Computer Music Association, pp. 531–534.

[96] Wassermann, K. C., Verschure, P., and Manzolli, J. Live soundscape composition based on synthetic emotions. *IEEE Multimedia 10*, 4 (2003), 82–90.

[97] Watson, K. B. The nature and measurement of musical meanings. *Psychological Monographs*, 54 (1942), 1–43.

[98] Wedin, L. Multidimensional study of perceptual-emotional qualities in music. *Journal of Psychology*, 13 (1972), 241–257.

[99] Wiggins, G., Papadopoulos, G., Phon-Amnuaisuk, S., and Tuson, A. Evolutionary methods for musical composition. *International Journal of Computing Anticipatory Systems* (1999).