I-Sounds

Emotion-based Music Composition for Virtual Environments

Ricardo Miguel Moreira da Cruz

Instituto Superior Técnico
Intelligent Agents and Synthetic Characters Group
Instituto de Engenharia de Sistemas e Computadores
Lisboa, Portugal

Abstract. 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

1 Introduction

Music is an art form consisting of sound and silence. This extraordinary manifestation of the human intellect has been a subject of great interest for Science and Philosophy. One interesting topic is that of 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. This work aims to provide new research tools and applications for the study of musical affects, exploring the frontier between music and our emotions.

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 but the ultimate and common goal is the

improvement of the user's or audience's experience while enjoying the interaction with such applications. Today, most applications use pre-composed soundtracks with a varying level of refinement, but while this approach offers some advantages such as, the high aesthetic value entailed by the composer's creativity, and the refined "illustrations" allowed by the script's careful study, it fails to provide the increasingly demanded flexibility; changes to the original script may enforce time and resource consuming changes to the soundtrack, and the user is more likely to loose his interest quickly while having a less immersive experience. Although automatic or semi-automatic music composition lacks the aesthetics of a real composer it is the best option to provide such flexibility.

The present work aims to develop a flexible solution for music composition from affective information. The idea is to provide a tool that can be easily integrated with interactive affective systems, and use the auditive interaction channel to enable a richer interaction. An architectural proposal for a development framework, as well as, a run-time environment to integrate those algorithms with any affective system, a full implementation of the proposal architecture, and the development of a concrete affective composition algorithm, are the three components that 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.

2 Proposed Architecture

This proposal together with its implementation 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, and thus, this proposal is rather flexible and extensible, postponing context dependent decisions.

Three high level requirements bounded the design of the proposed architecture; generality and extensibility, real-time performance, and high integrability. There is an enormous variety of affective systems, using different conceptualisations, e.g., emotion models, thus, a general architecture should be flexible enough to accommodate each approach's specificities. 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. Considering the myriad of development environments and tools, as well as, techniques, it is important to provide a flexible integration mechanism, whose challenge is to define a common model that can account for various integration options while ensuring a detail independent and coherent assembly.

2.1 The Run-time Environment

Composition systems can be structured in terms of three different stages; sensing, processing and response [3,5]. This architecture follows this paradigm, consisting of three processing layers; affective, composition and output, that resemble the sensing, processing and response stages respectively. This approach offers important advantages; it reduces implementation complexity, and since each module is a rather independent piece its development and extension does not depends on the other two.

The affective layer manages an intermediate representation of the affective application state – it can be seen as a big affective buffer. This layer maintains an affective environment, consisting of affective entities each one with its own emotional state. The composition layer uses that affective data to compose real-time music. Music composition is done through an asynchronous composition pipeline, consisting of processing stages that can be connected at will by the programmer. The output layer is responsible for output operations, implemented in the form of output handlers, that encapsulate the necessary logic for data output in the desired channel. Multiple simultaneous outputs are possible.

2.2 Application Programming Interface

In addition to the run-time environment, the proposed architecture encompasses an application programming interface that can be used to extend the systems functionalities and models, and develop composition algorithms.

Most of the interaction with the affective layer is done through the affective environment. The proposed architecture is neutral in what respects to emotion models, the programmer is free to implement the most appropriate models for his application. Interaction with the composition layer is done through the composition pipeline component. Specific composition algorithms are implemented in the concept of Composer, which is a set of connected processing stages that can be loaded and unloaded at request. Interaction with the output layer is done exclusively through the output manager, and the API provides the necessary functionality to implement specific output handlers.

2.3 Integration Driver

The proposed integration model is based in the concept of driver. A driver is a small program working between the affective application and the composition system, receiving information and control requests from the affective application and executing the adequate actions over the composition system through the proposed API.

2.4 Event Notification and Logging

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. On the other side,

such notification mechanisms can be used to develop useful tools able to interact with the composition system such as, graphic user interfaces.

Usually it is useful to analyse the system's behaviour, during or 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. This proposal introduces an optional and flexible logging facility, that is also available, through the system's API, to the developer using the I-Sounds framework as a platform for his composition algorithms.

3 AMADEUS

AMADEUS is a composition algorithm, named after the eighteenth century composer, Wolfgang Amadeus Mozart, and implemented on top of the I-Sounds framework, which is able to compose short music segments expressive of certain emotions, those provided as the algorithm's input. From an affective point of view, AMADEUS development was centred around a reduced set of emotions; happiness, sadness, anger and fear. Happiness and sadness, are two popular choices among researchers, the main reason for such popularity might be related with their bipolar character, and due to this fact they are easier to express and distinguish. 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's expressivity.

Music is the product of many different structural factors and 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. Building a composition algorithm that uses a relative high number of structural factors at once, is therefore an 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. Such approach, should focus initially, 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; what are those fundamental base factors?

3.1 Affective Rhythm

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 note and rest, 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 note emerges rhythm, transformed in meter when subject to a rule, a measure usually denoted by a time signature. 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, but historically its role was not always recognised as such.

The composition algorithm proposed in this work, AMADEUS, is for the most part based on rhythm, which incidentally, is superficially studied - surely this is related with its complexity. To develop AMADEUS it was clear the need for a deeper and solid perspective on rhythm that somehow contemplates its perceptive qualities. Such perspective is naturally contributed by a percussionist; in his PhD thesis, Eduardo Lopes [2], develops an extensive theory of rhythm and meter with adequate empirical support. The concepts of pulse salience and kinesis introduced by Lopes as the two fundamental perceptive qualities of rhythm, provided the necessary theoretic foundations upon which the rhythm composition module of AMADEUS can be based, however, this proposed two challenges; to provide the necessary formalisations and quantification mechanisms for pulse salience, and use an essentially analytical theory as a composition tool. AMADEUS addresses the second question with a two step strategy; the first step was to develop appropriate measures, while the theoretical background states that those 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 [4]. By combining this mapping information with the original Russell's model, hybrid maps of emotions and rhythm perceptive qualities can be built, from there composition rules can be inferred and programmed in AMADEUS.

Pulse salience. Pulse salience is 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.

Metric position is the salience component depending on the metric placement of each pulse. 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 organised by listeners. 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. The component of agogic accentuation depends exclusively on the pulse's duration; the longer the pulse the more stable it is perceived. For quantification purposes, each note type, e.g., whole notes, half notes and eighth notes, are assigned a number $n \in N$. The assigned 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. The overall salience of a pulse is also 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. This component is referred to as rhythm cell accentuation, because 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. Pulse salience is therefore, a first order polynomial function, given by (2).

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.

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 (1).

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

$$S(\eta, \omega) = (BU - M\omega) + \eta + \sum_{(\zeta,\sigma)\in C} \left(\left\lceil R\left(\frac{\eta}{\zeta}\right) M\sigma \right\rceil \right) \tag{2}$$

then pulse salience can be calculated by (2) being;

η is the pulse value.

ω is the pulse metric value.

Kinesis. Kinesis refers to the perceptive quality of rhythm responsible for the music's 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. Tension can be 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.

3.2 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. The major mode is usually classified as happy and joyful, while the minor mode is usually classified as sad and melancholic. In what respects to melody and the diatonic modes, AMADEUS uses only the three fundamental notes necessary to define a major or minor context, the tonic, the mediant, and the dominant, e.g., in a C major context the notes used would be C, E and G. 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, starts each single measure segment with a C and terminates with a E.

4 Evaluation

This section 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 [1], 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.

4.1 AMADEUS calibration

AMADEUS uses a set of intervals to constraint the allowed salience value for each metric position. In this first experiment, AMADEUS was parametrised with four interval sets, one for each expressed emotion; happiness, sadness, anger and fear. These candidate intervals were based in the author's empirical music knowledge. The experiment consisted in the evaluation of 3 segments for each emotion for a total of 12 segments, composed in quaternary meter, denoted by a 4/4 time signature and played with a constant tempo of 80 beats per minute. Each segment varied in the rhythmic constructs and mode, C major or C minor. A total of 68 volunteers, 53 males and 15 females, classified all of the twelve segments.

The experiment results were very encouraging for the expression of happiness, the participants were able to clearly distinguish the happy segments from the others, and a correlational analysis suggests that happiness is the easiest emotion to communicate through rhythm and mode. In what respects to the perception of sad segments, good results were also achieved but they also suggest that a refinement of the salience intervals constraining the composition of those segments might improve its recognition. 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. A differences analysis was conclusive in what respects to classification differences between genders and the participants music knowledge, no significant differences were registered for these two factors.

4.2 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, is 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 group of 24 Portuguese boys and girls, in the scholar age between 8 and 9 years old participated in the test. The experiment consisted in the projection of four video-clips, two times each, with and without music accompaniment, to two different test groups. Each clip depicted a character, exhibiting a certain motion pattern expressive of the four possible emotions. After the projection of each clip, the participants were asked to immediately evaluate the characters emotion, using an evaluation matrix.

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. The recognition of sadness, was also clear; an improvement of 30% was registered between the condition of no music accompaniment and the one where music accompaniment was used. The results for the other two emotions are less clear; in absolute terms the recognition of anger suffered a slight 5% decrease from the condition with no music accompaniment, contrasting with 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, which can also be circumstantial, but on the contrary it represents an improvement. 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, 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.

4.3 Story Intelligibility

The measurement of the emotion recognition rate with isolated clips featuring a single character provides good experimental control, and it is an adequate approach to quantitatively assess music's contribution. However, it is insufficient, to assess the overall qualitative improvement of the users experience and the

overall narrative intelligibility. This experiment, more concerned with ecological validity, addresses this later situation. The experiment's goal was to qualitatively assess the overall improvement of the user's interaction. Each participant was asked to answer a small questionnaire about the emotions expressed by the characters. 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. A group of 24 Portuguese boys and girls, in the scholar age between 8 and 9 years old.

The qualitative data collected during the experiment is not suitable 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 with 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. The results together with the experimenters observations and participant's 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.

5 Conclusions

The I-Sounds project is a practical application of musical emotions to improve the user's interactive experience in affective systems. Although, it goes a step ahead and proposes a new perspective, over one of the least studied music structural factors, rhythm. One of the objectives of this work was the definition of suitable development tools for emotion-based composition algorithms. The proposed architecture and its implementation, provides such a suitable framework to develop and explore the composition of affective music. Far from being limited to algorithm development, this framework, provides the necessary run-time components to integrate and run those algorithms with affective systems, but 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. 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, 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 evaluated the success rate in the recognition of the emotions exhibited by an animated character, and measured the improvement in the user's perception and intelligibility of a full narrative performed by I-Shadows. The results emerging from these experiments are encouraging. The usage of music had greatly improved the recognition rate of happiness and sadness. 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. The experiments had positively assessed the influence of music, in the general improvement

of the user's experience with I-Shadows. The young participants, with an age ranging between 8-9, demonstrated a greater interest and commitment with the tests when music accompaniment was used.

5.1 Future Work

Rather than a complete solution, this work is above all, a proof of concept. To accomplish its vision, the 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 remains unexplored. 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. However, far from being an exclusive result of rhythm and mode, music has many more factors that can be considered. Irrespective from the future researcher's choice, a bigger concern and care should be put in the control of multiple factor interactions. 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.

References

1. Brisson, A., Fernandes, M., and Paiva, A. Children as affective designers. In *HUMAINE WP9 workshop* (Kista, Sweden, 2006).
2. Lopes, E. *Just In Time: Towards a theory of rhythm and metre.* PhD thesis, University of Southampton, UK, 2003.
3. Rowe, R. *Interactive music systems: machine listening and composing.* MIT Press, Cambridge, MA, USA, 1992.
4. Russel, J. A. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39 (1980), 1161–1178.
5. 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).