The Music and Emotion Driven Game Engine: Ideas and Games

Roberto Dillon, Ng Kian Bee and Simon Rozner

School of Interactive and Digital Media - Nanyang Polytechnic
180 Ang Mo Kio Av.8
Singapore 569830

Abstract

In this paper we describe the ideas behind the Music and Emotion Driven Game Engine (M-EDGE), currently under development at the School of Interactive and Digital Media in Nanyang Polytechnic and fully supported by the Singapore National Research Foundation. The paper will explain a possible method for analyzing emotional content in music in real time and how it can successfully be applied to different game ideas to help defining a new interactive experience and music based gameplay in videogames.

Keywords

Music based games, game API, cognitive musicology, real time human-computer interaction, usability.

1. Introduction

Music is commonly used in a variety of media, like movies, theatric plays and games, for defining and emphasizing the emotional impact of the show while immersing the spectators in the described plot. The audiences, or players in the case of games, are, anyway, tied to a passive role by listening to how the system uses music to define the emotional characteristic of the different scenes.

In the M-EDGE project, instead, we aim at building an innovative system that pushes this boundary forward by allowing players to be actively involved in communicating their own expressive intentions and moods to the computer through music by playing a real musical instrument of their choice in real time. Music can then by used as a controller to affect the events in the virtual world and, ultimately, play the games. To achieve this, we first have to define a method for extracting and classifying emotional content in music and then implementing it in an easy to use API that can be integrated in different game engines to develop a new breed of games based on this approach.

2. Analyzing emotional content in Music

Analyzing emotional content in music is not a new topic in the musicological and cognitive science communities and both theoretical and experimental approaches have been proposed in the past. See, for example, [1], [2], [3] and [4]. In particular, the idea of squaring and low pass filtering the musical input with filters having different cut off frequencies to extract two profiles related to envelope and energy of the signal and then comparing them for extracting a series of audio features (figure 1), was first proposed in [5] and then successfully implemented in different experiments including [6], [7] and

[8] where the two profiles are extracted by using cut off frequencies around 20Hz and 1Hz respectively. Since this approach proved to be reliable in the past, we decided to base our research and build our system upon it too.

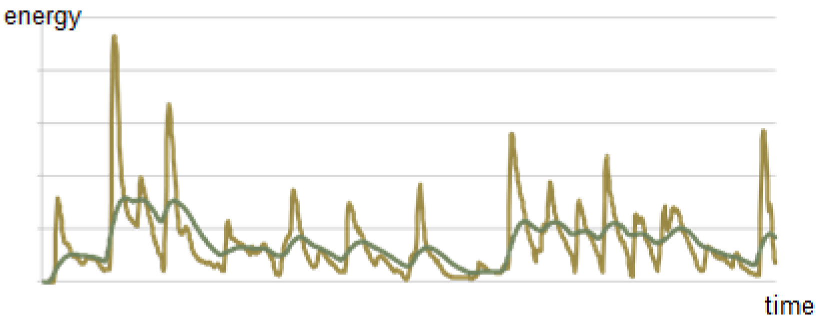

Figure 1. 20Hz and 1Hz profiles extracted by squaring and low pass filtering an audio file to divide it in a series of events (an event starts when the 20 Hz profile goes above the 1Hz profile and ends when it goes below it). These can then be used to extract a set of different audio features, also known in literature as “audio cues”, which can be related to expressive intentions and emotional contents.

3. Choosing the most suitable Audio Features

Different audio features may be more appropriate than others according to different settings, musical instruments and playing styles, so it is important to have a system able to extract as many features as possible and then choose the most effective ones on a case by case basis. In particular, the M-EDGE API is able to extract mean and standard deviation of a wide variety of features related to tempo, dynamics and articulation of the music by analyzing each individual event across a predefined time window, besides also extracting energy levels across different frequency bands.

To have an insight on which features might be most appropriate for our games, we arranged a recording and listening experiment to build a database of short musical samples which would work as a test bed and reference library for our system.

3.1 Setting up the experiment

The first stage of our experiment consisted in inviting twenty musicians of different skill levels, from amateurs with as little as one year of musical practice to highly skilled professionals, performing on different instruments (flute, drums, violin and acoustic guitar). They were asked to improvise three short tunes to represent each of the following basic emotions: Anger, Happiness, Sadness and Fear. The performances were recorded at 44.1KHz and 16 bits and made up the reference database needed for further study.

3.2 Web based test

Once the recordings were done, we arranged a web based test to evaluate which ones, among all the files in our database, were most effective in expressing the particular emotions. To do this, we designed a simple webpage and asked seventy people (most of them NYP students with little or no musical background) to listen to the samples in random order and identify which basic emotion the player was trying to express. Besides the four nominal emotions, they were also allowed to select an option stating they were not able to identify any particular emotion clearly.

3.3 Data Analysis

Results from the web test were rather surprising: while we were able to identify performances on all instruments that got a high percentage of correct recognition (more or equal than 85%) for Happiness, Sadness and Anger, none of the recordings portraying Fear obtained such good ratings with the notable exception of performances from one of the professional violinists (a first part from Singapore Symphony Orchestra) who did extensive use of advanced techniques to simulate a variety of effects. In general, we realized that “Fear” was often misunderstood as “Sadness”, if played in a very ambiguous and shy way, or as “Anger” if played trying to mimic panic (see Figure 2).
In the end, since we wanted to model players’ styles as accessible as possible, including amateurs and not so experienced people, we decided to focus only on the three basic emotions that showed to be more predictable and easier to identify by a larger group of people, so “Fear” was excluded from the rest of the study.

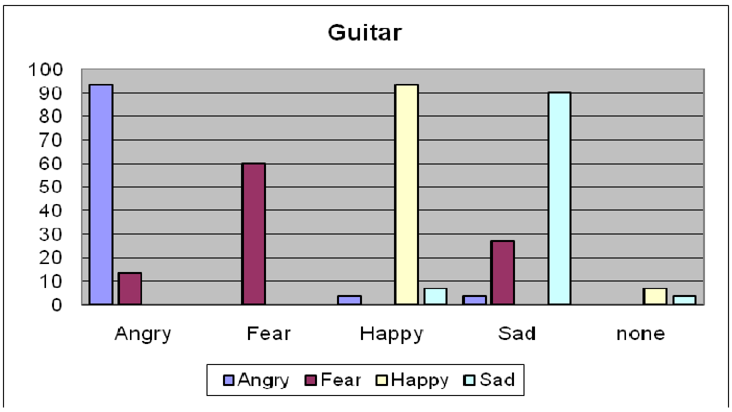

Figure 2. Recognition statistics for the performances of one of the guitarists who took part in our experiment. As we can see, Angry, Happy and Sad were properly recognized at least 90% of the time while Fear barely reached 60%.

The next step was to evaluate how well each of the possible audio features can identify the differences amongst the performances and then identify the two or three best ones upon which building a reliable classification system.
To achieve this, for each instrument, we did a single factor ANOVA comparing each feature across the three different emotions and then, to get a better insight, a t-test assuming unequal variances was also performed on each feature comparing each emotion with one of the others. It turned out that, in general, the three best features were the mean values of the Number of Events per Second, Onset Level (defined as the value of the 20Hz profile when it intersects the 1Hz profile) and the Attack Velocity (defined as the first derivative of the 20Hz profile at the onset point) with two of these being more relevant in different instruments. For example, ANOVA and t-test results for the guitar

performances that scored the highest recognition percentages are shown in Tables 1 through 4 where the two most meaningful features turned out to be Attack Velocity and Onset Level.

Groups	Count	Sum	Average	Variance		
angry	252	12.39625	0.049191468	0.000464		
happy	294	4.084219	0.013891901	3.95E-05		
sad	323	1.247126	0.003861071	2.46E-06		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.3097	2	0.154849856	1041.628	3.6E-231	3.00612
Within Groups	0.128741	866	0.000148661			
Total	0.43844	868				

Table 1. Single factor ANOVA results for the Attack Velocity feature (Guitar). α = 0.05

	angry	happy		angry	sad		happy	sad
Mean	0.0491915	0.0138919		0.0491915	0.0038611		0.0138919	0.0038611
Variance	0.0004636	3.951E-05		0.0004636	2.462E-06		3.951E-05	2.462E-06
Observations	252	294		252	323		294	323
Hypothesized Mean Difference	0			0			0	
df	288			253			326	
t Stat	25.123148			33.350672			26.618568	
P(T<=t) one-tail	7.501E-75			6.804E-95			4.739E-84	
t Critical one-tail	1.6501617			1.6508987			1.6495412	
P(T<=t) two-tail	1.5E-74			1.361E-94			9.477E-84	
t Critical two-tail	1.9682351			1.9693847			1.9672675	

Table 2: results for t-test, two samples assuming unequal variances for the Attack velocity feature (Guitar) across the different performances.

Groups	Count	Sum	Average	Variance		
angry	252	42.40633	0.168279	0.003833		
happy	294	18.13893	0.061697	0.000441		
sad	323	5.24581	0.016241	3.48E-05		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3.359441	2	1.679721	1319.339	1.3E-263	3.006119
Within Groups	1.10255	866	0.001273			
Total	4.461992	868				

Table 3. Single factor ANOVA results for the Onset Level feature (Guitar). α = 0.05

	angry	happy	angry	sad	happy	sad
Mean	0.1682791	0.061697	0.1682791	0.0162409	0.061697	0.0162409
Variance	0.0038334	0.0004408	0.0038334	3.479E-05	0.0004408	3.479E-05
Observations	252	294	252	323	294	323
Hypothesized Mean Difference	0		0		0	
df	300		255		335	
t Stat	26.072237		38.844484		35.855988	
P(T<=t) one-tail	2.192E-79		2.19E-109		5.13E-117	
t Critical one-tail	1.6499487		1.6508511		1.6494149	
P(T<=t) two-tail	4.384E-79		4.37E-109		1.03E-116	
t Critical two-tail	1.9679029		1.9693105		1.9670705	

Table 4: results for t-test, two samples assuming unequal variances for the Onset Level feature (Guitar) across the different performances.

By using these two features it is possible to visualize in a simple 2D space how the data cluster according to the different emotions (Figure 3), showing that it is possible to build a system able to discriminate between them.

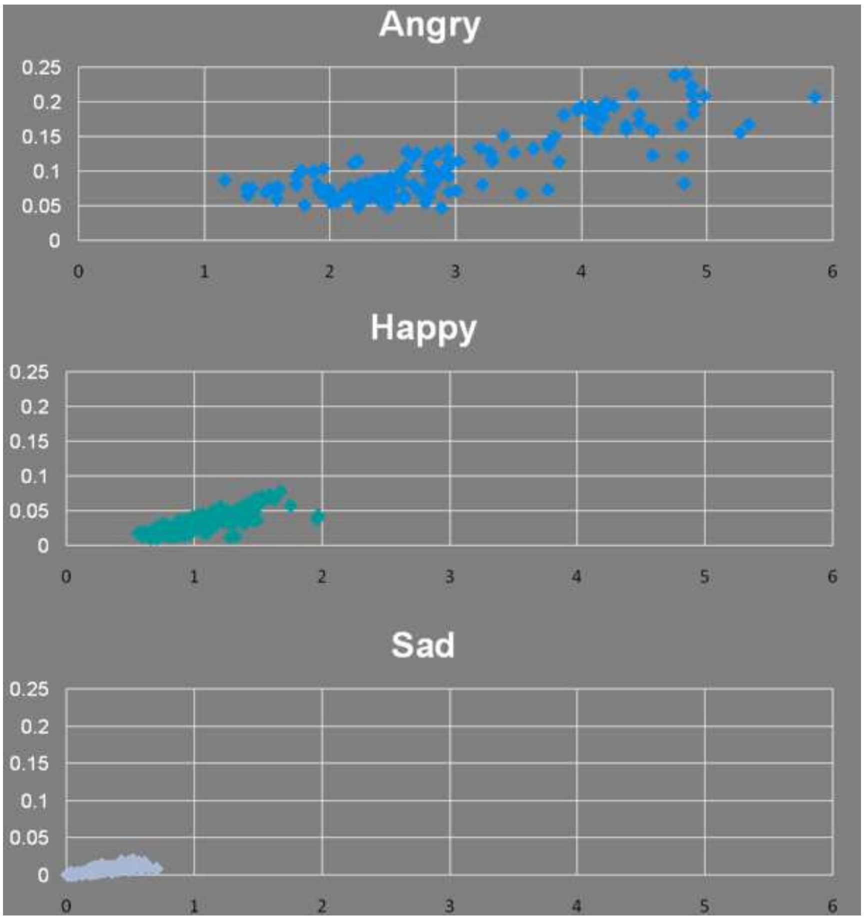

Figure 3. plotting the data extracted from the three different guitar performances. Onset Level on the X axis, Attack Velocity on the Y axis. Onset level data are normalized to the mean value of the Happy performance.

4. Recognition System

Once verified that the chosen features show statistically meaningful differences between the performances, many different approaches can be used for building an automatic classification system. In our case, we wanted something robust but also as simple as possible to be implemented in our API so as to save most of the computational resources for other game related tasks.

In particular, note that we have assumed the distribution of the extracted features to be Gaussian thanks to the Central Limit Theorem so, once the system extracts the different features, we can simply compute the probability of a value X belonging to a particular distribution (i.e. emotion) by:

$$P(X) = \frac{1}{\sqrt{2\pi\sigma}} \int_{X-\delta}^{X+\delta} e^{-(x-\mu)^2 / 2\sigma^2} \, dx$$

Where μ and σ and the mean and standard deviation of the particular distribution and δ is defined as σ/10.

5. The M-EDGE API

All the functions to record audio, extract all the features and compute the different probabilities, given a particular μ and σ that can be computed by asking the player to provide reference samples of his own playing style, were implemented in a lightweight C++ library that can be integrated into different game engines like C4 [9] and which forms the core of the M-EDGE system.

Besides this, since M-EDGE is aimed at developing games that are likely to be played in many different circumstances and settings by many different people, we realized that different features may be more relevant than others for different players while others may be completely misleading and unable to discriminate between the performances. This means that constraining the system to use only the default features that worked in our experiment may not be a wise choice so, to make the system more robust and reliable, also a simple function named “AWE” (i.e. Automatic WEighting) was implemented. This function analyzes the reference distributions of the extracted features in the player’s performances and gives higher priority to those showing bigger differences between the various emotions.

6. Games

We believe that M-EDGE can be a viable tool to help designing many different games, ranging from educational music based software to innovative minigames in more complex and standard products. Besides these, though, also more serious topics can be addressed like rehabilitation games useful and accessible to different people who may have physical/rehabilitation problems. It may also serve as a tool to assist physicians and patients with particular forms of cognitive/emotional impairment, due to brain damage in the frontal lobes area, in their difficult path to understanding and communicating emotions again.

So far, to test the M-EDGE system’s capabilities, a couple of simple casual games were developed, Moody Balloons and Vocal Edge, and a more ambitious project, Garden Planet, is currently in production. Available games can be freely downloaded from the M-EDGE website [10].

6.1 Moody Balloons

Moody Balloons (Figure 4) is a very simple casual game where the player has to pop as many balloons as possible before they fill up the screen. Each balloon is characterized by one of the three basic emotions and will pop only if the player is able to perform the corresponding emotion long enough by improvising new tunes and rhythms.

Figure 4. Moody Ballons, the first M-EDGE based game

6.2 Vocal Edge

Vocal Edge (Figure 5) is a simple singing game where, besides trying to hit the correct notes, the player is also required to interpret the songs correctly by showcasing the proper emotions.

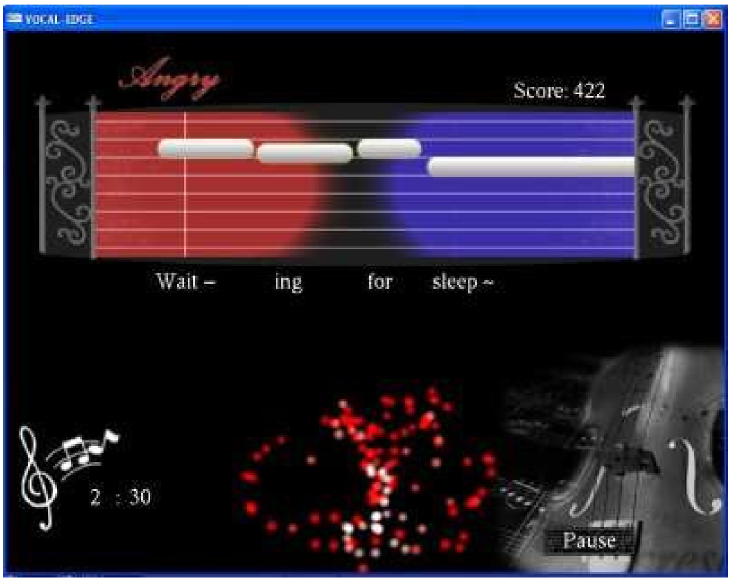

Figure 5. Vocal Edge

6.3 Garden Planet

Garden Planet is a 3D strategy game where the goal is to populate a barren planet with plant life and grow a beautiful garden. The user has control over the elements in order to spread plant seeds, help plants grow and manipulate the land to promote a healthy environment (Figure 6).

The game tries to simulate a planet that is directly impacted by the user's actions and choices. To achieve this, M-EDGE is used as a link for the input by the user to control the elemental powers (i.e. wind, water and earth plus fire which is controlled indirectly by the other three) where the music the user is playing controls their strength.

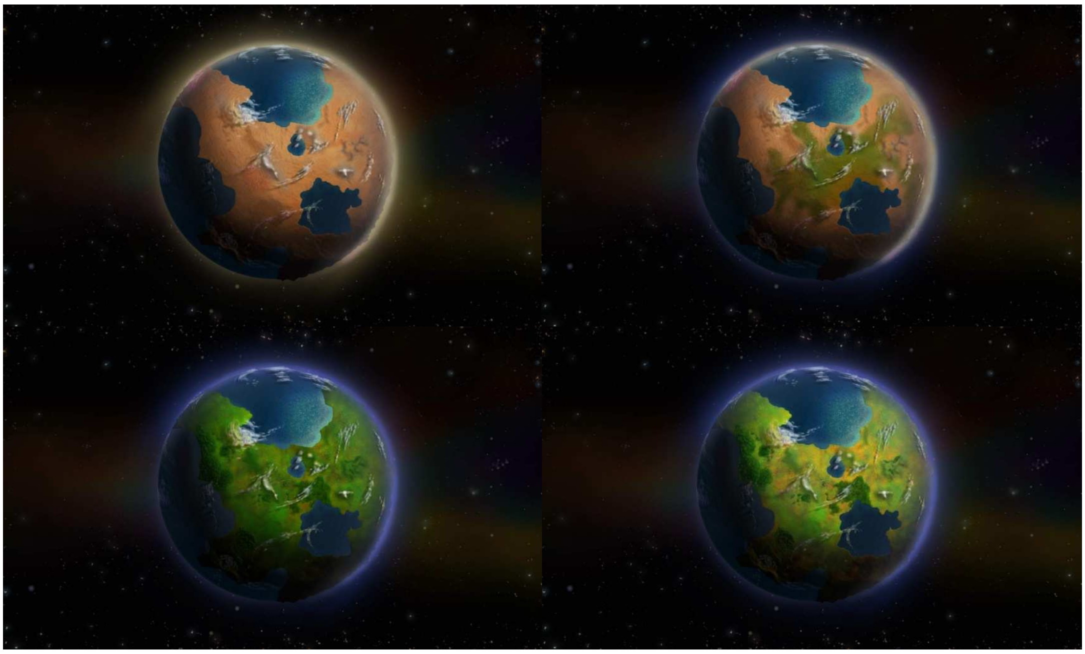

Figure 6. Garden Planet: From top left, clockwise: shaping a barren planet from a desert into a luscious garden.

Besides this, the music causes the planet to have an emotional reaction on its own as well. For example if mainly angry music is played the planet might get aggravated by this anger. It shows it to the player by changing the general color scheme and causing more natural disasters that can threaten the existing plant life. This "emotional" response of the planet has in return a direct impact on the user and how he will interact further with the game.

Overall, Garden Planet stresses the emotional aspect of music and how it is perceived by someone else then the person playing it. It shows the user how it "feels". Garden Planet is a game with an ongoing conversation of emotions using music as the input device for the user.

Conclusion

Current results and the interested feedback we obtained so far confirm that M-EDGE can be an innovative tool in the hands of resourceful game designers: the possibility of enriching players experiences by considering music and its emotional contents as an actual input to control gameplay

mechanics can be efficiently implemented and can add a new layer of depth to games targeted towards not only musicians but also to a more general audience. Our first experiments, in fact, showed that, even though these games seem more suitable to people with some musical background who can improvise new tunes on the fly, even people with little or no musical experience can enjoy them by having fun experimenting with simple drums or just by singing and humming different tunes.
More experiments are on the way and games like "Garden Planet" may provide hints on the viability of this kind of games also at a commercial level.

Acknowledgments

M-EDGE is supported by the Singapore National Research Foundation under the Interactive Digital Media R&D Program, research grant NRF2007IDM-IDM002-015.
Special thanks also to all NYP management and SIDM staff involved (Mr. Ron Ang, Mr. Sridhar Annalamai, Mr. Chen Wei Ren, Dr. Guo Zheng, Mr. Yao Kheng, Mr. Andrew Lam, Dr. Li Jiang, Mr. Low Cheng Lai, Mr. Alex Toh, Mr. Yin Xiang) besides all the students who gave continuous feedback and are collaborating on different aspects of the project, including the development of the various games. A detailed list of team members and roles is available at http://www.GamesCreation.org/medge

References

[1] Cooke D. (1959): The Language of Music, Oxford University Press
[2] Sloboda J. (1991): Music structure and emotional response: Some empirical findings. Psychology of Music, vol. 19, pp.110-120
[3] Sundberg J., Friberg A., Fryden L. (1991): Common secrets of musicians and listeners: an analysis-by-synthesis study of musical performance. In Representing Musical structure, Academic Press
[4] Gabrielsson A. (1995): Expressive intention and performance. Music, Mind and Machine, New York, Springer, pp.35-47
[5] Todd N.(1992): Music and Motion: a personal view. In Proceedings 4th Workshop on Rhythm Perception, Bourges, France
[6] Dillon R. (2001): Extracting audio cues in real time to understand musical expressiveness. In Proceedings "Current research directions in computer music", MOSART Workshop, Barcelona, Spain, pp.41-44
[7] Dillon R. (2003): A statistical approach to expressive intention recognition in violin performances. Proceedings SMAC 03, Stockholm, Sweden, pp. 529-532
[8] Dillon R. (2003): Classifying musical performance by statistical analysis of audio cues, Journal of New Music Research, Vol. 32 n.3, pp.327-332
[9] C4 Game Engine: http://www.terathon.com
[10] M-EDGE Website: http://www.GamesCreation.org/medge