DYNAMIC GAME SOUNDTRACK GENERATION IN RESPONSE TO A CONTINUOUSLY VARYING EMOTIONAL TRAJECTORY

DUNCAN WILLIAMS[1] ALEXIS KIRKE[1] JOEL EATON[1] EDUARDO MIRANDA[1]
IAN DALY[2] JAMES HALLOWELL[2] ETIENNE ROESCH[2] FAUSTINA HWANG[2] SLAWOMIR NASUTO[2]

[1] *Interdisciplinary Centre for Computer Music Research, Plymouth University, UK*
duncan.williams@plymouth.ac.uk
[2] *Brain Embodiment Laboratory, University of Reading, UK*

Dynamic soundtrack creation presents various practical and aesthetic challenges to composers working with games. This paper presents an implementation of a system addressing some of these challenges with an affectively-driven music generation algorithm based on a second order Markov-model. The system can respond in real-time to emotional trajectories derived from 2-dimensions of affect on the circumplex model (arousal and valence), which are mapped to five musical parameters. A transition matrix is employed to vary the generated output in continuous response to the affective state intended by the gameplay.

1 INTRODUCTION

Musical scores have a documented effect on the overall user experience of a game, including heightened immersion and engagement [1]. The world of sound-to-picture in film and television has long been able to exploit these tendencies, in order to directly heighten affective engagement in audiences [2] but the nonlinear nature of gaming makes traditional soundtrack creation a more difficult proposition. Player immersion can be destroyed by repetition, which in itself is much more likely when playing a game than, for example, watching a film (in which repetition can be expressly controlled to the advantage of a fixed timescale). Various strategies for tackling this problem by adapting the soundtrack to gameplay exist [3]. Sequencing pre-composed (and typically, pre-recorded) sections of music dependent on gameplay is one such strategy. However, the illusion of a real-time, responsive soundtrack then becomes dependent on the 'resolution' of the transitions between these sections. Longer transitions and loop points make for less convincing, and less immediately responsive, soundtracks, whilst the use of shorter transitions and a correspondingly larger number of sequences makes for a great deal of compositional complexity in the development of the score. Branching strategies offer a prescribed number of routes through a soundtrack, whereby breakpoints are used to indicate progression to another branch [4]. Visual transitions can be accompanied by audible cross-fades in the case of reaching a gameplay breakpoint before the end of a sequence, and codas abruptly inserted to accompany the end of play [5]. But in either case, the risk of repetitive soundtrack creation if a player stays within the boundaries of a single branch for an extended period of time remains. One alternative is to continuously generate new music using algorithmic composition. Many generative models for algorithmic composition exist and the field is now reaching a degree of maturity [6]–[8]. However, the use of algorithmic composition techniques to generate music with emotional meaning, such as might be best used to augment narrative in games, is still in its infancy [9], [10]. This paper presents the implementation of one such affectively-driven algorithmic composition system which responds to emotional trajectories in 2-dimensions to continuously create novel music in real-time.

2 PREVIOUS WORK

High quality soundtrack creation can enhance multimodal engagement (outwith music, both foley and automated dialogue replacement have been used to great success in the sound-for-picture arena). Gaming, however, presents a different challenge to film soundtrack creation in that the timeline is almost always nonlinear, with varying degrees of player control involved. The nonlinear time frame required by most gaming soundtracks has been tackled with various solutions, of which perhaps the most commonly used is to loop a given section of music until a breakpoint (for example, a change of scene, the end of a level, death of player character etc) is reached. This type of system is illustrated in Figure 1.

The disadvantage of such an approach is that the repetition involved is high and can thus begin to hinder player immersion, becoming distracting or even outright irritating at transition points if they are not managed carefully. The effective resolution of such systems can be improved by adding divergent 'branches' from breakpoints in the musical sequence. The resulting soundtrack is less linear, but the complexity involved is increased by the need to create a series of score fragments which can be interchanged whilst maintaining the intended aesthetic (remembering that the purpose of augmenting games with high quality soundtrack creation is to increase player immersion and emotional engagement, and that composing emotionally engaging music is in itself a non-trivial task).

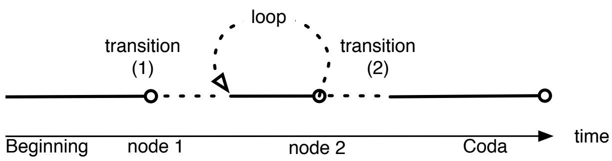

Figure 1. Adapting a soundtrack to three narrative sections with some control over time. Players progress to the middle section, which plays on a loop until the next breakpoint is reached. Transitions between sequences must be precomposed. This approach has been widely used in gaming (see, for example, the *Final Fantasy* series).

2.1 Affective models and branching strategies

Music psychology often documents emotion, affect, and mood as the three types of emotional responses listeners might exhibit in relation to music, each increasing in duration [11]. Russell's circumplex model of affect [12] provides a way of parameterising emotional responses to musical stimuli across two dimensions: *valence* (a scale of positivity) and *arousal* (a scale of energy or activation strength), as shown in Figure 2. Emotional descriptors from Hevner's adjective cycle can be mapped quite closely onto this model [13] and have been corroborated by other studies of emotional responses to music in 2-dimensions [14, p. 22]. Interested readers can find more exhaustive reviews on the link between music and emotion and the various types of emotional models employed in [15] and the recent special issue of *Musicae Scientiae* [16].

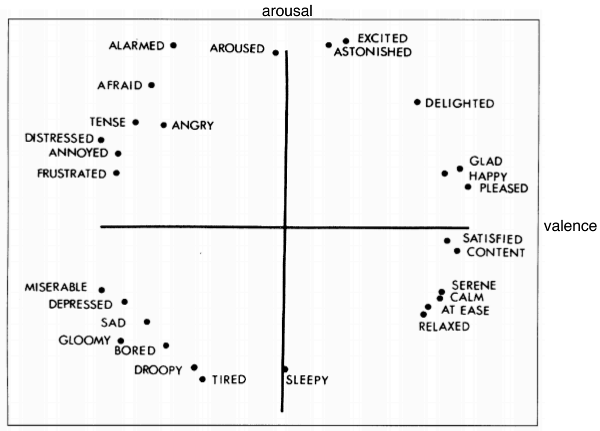

Figure 2. Circumplex model of affect, from Russell, p1168 [12], with addition of valence and arousal labels at each axes (author's own emphasis) Adjectives have been scaled in two dimensions, with valence on the horizontal axis and arousal on the vertical axis.

Targeting emotional responses with affectively-driven soundtrack creation has obvious advantages when accompanying gaming narrative and enhancing player immersion. However, the adaptability of a branching system to emotional responses in 2D is minimal, due to the complexity involved in creating an appropriate number of pre-composed branches. Figure 3 shows a basic branching system with two decision points (at nodes 2 and 3). In terms of gameplay this can represent the simplest type of score that might follow a narrative: beginning, middle, and end. Note that 7 discrete sequences of music would need to be composed and recorded, with appropriate musical transitions at each node, before being loaded into RAM to be triggered by gameplay at each decision point.

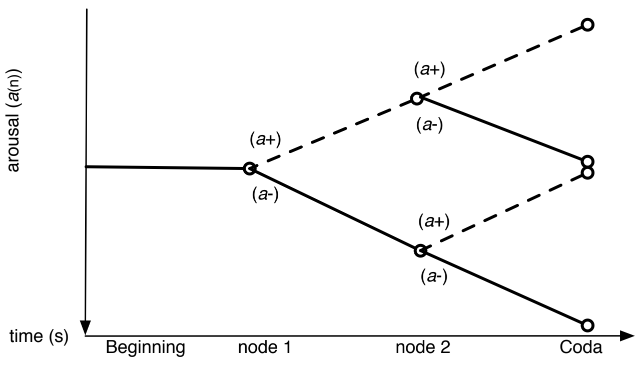

Figure 3. Branching system used to sequence a nonlinear musical score based on *arousal* cues in the simplest form of game narrative (beginning, middle, end). Nodes represent decision breakpoints based on gameplay. (*a*+ and the dashed line indicates high arousal at and after node, *a*- and straight line

indicates low arousal). 7 discrete pieces of music are required by this approach.

Figure 4 shows the additional complexity that simply adding 'high' or 'low' values for valence to each node would cause – 21 separate sequences of music are now required to present just four emotional states (arousal high/valence low, arousal high/valence high, arousal low/valence low, and arousal low/valence high). To represent the 12 affective states shown in Table 1, over three narrative stages, this type of branching system would need to include (at least 61) discrete musical sequences with appropriate transitions – additional decision points (nodes) would increase the amount of compositional complexity exponentially.

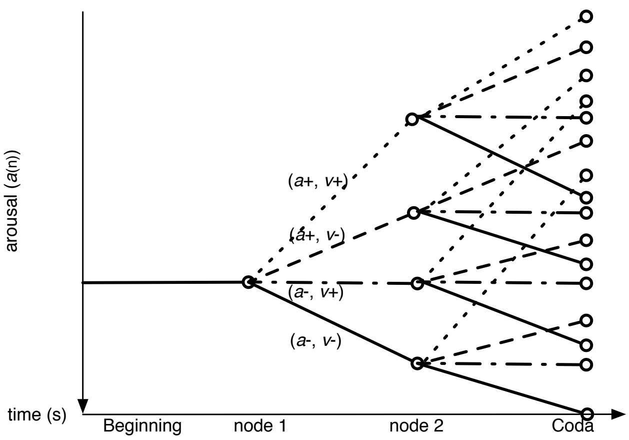

Figure 4. Branching system used to sequence musical score based on *arousal* and *valence* cues in the simplest form of game narrative (beginning, middle, end). Nodes represent decision breakpoints based on gameplay. (*a*+, *v*+ and the dashed line indicates high arousal and high valence, *a*+, *v*- and the dashed line indicates high arousal, low valence, *a*-,*v*+ and the dot-dashed line indicate low arousal, high valence, and *a*-, *v*- and the solid line indicate low arousal and low valence). 21 discrete pieces of music are required by this approach.

Thus the choice of branch in such systems can be made dynamically and adaptively in direct response to gameplay, aiming for a richer, more immersive experience than a linear system. However, the larger numbers of branches involved equate to a hugely increased level of compositional complexity, as each branch must still be pre-composed, and each possible route auditioned for musical coherence before it can be committed to for use in the finished product. Furthermore, restrictions on computational processing overhead have been common to many game soundtracks in the past, and the increased number of nodes and subsequent musical segments required to achieve a reasonable 'resolution' would place a significant increase on processing load. Algorithmic generation of the score 'on-the-fly' can overcome this restriction, but systems for algorithmic composition often fail to specifically target emotional responses [17].

3 PILOT SYSTEM IMPLEMENTATION

A system for specific targeting of affective responses by means of algorithmic composition is proposed, based on an initial evaluation which suggested that such a system could achieve perceptual isolation in affectively-driven automated composition [18]. We consider this system to be 'proposed' as the implementation has yet to be the subject of a perceptual affective evaluation. The pilot system uses a three-stage process to learn, generate, and transform new musical material created by probabilistically analysing seed pieces of pre-composed music.

A 2nd order Markov chain [19] is used to evaluate and then create musical sequences from the source material. The system learns musical structures (or more specifically, a symbolic representation of musical structures in MIDI format), and creates a probability vector for each of the values in the input material as blocks of notes (*states*) containing pitch and rhythm data. A transition probability matrix is then derived between these states which is employed at the generation stage of the process, in which new musical sequences are created according to the likelihood of, for example, a C# 1/8th note following a C 1/16th note and so on. Thus the system can create related music with almost infinite amounts of variation from just a small range of seed material. This generated material is then transformed in the required musical parameters according to a selected affective mapping.

3.1 Affective mapping

This system divides the circumplex model into quadrants that are indexed according to Cartesian co-ordinate values, giving 12 unique affective states, as shown in Figure 5. 12 adjectives are mapped to these states such that 'basic' emotions (*sad*, *calm*, *angry*, and *happy*), which have been well-documented in many music and emotion studies, are represented in each of the quadrants, with suggested descriptors for lower and higher arousal levels spaced as evenly as possible from these adjectives in each subsection. In this manner, a co-ordinate of (*v1, a1*) would refer to *tired*. Two adjectives were deliberately avoided in the selection process: *Sleepy* and *aroused*, in order to minimise adjectives with a duality of meaning. Both are positioned with approximately neutral valence in the circumplex model of affect (shown in Figure 2). *Sleepy* might be considered by some to be a physical rather than an emotional response. *Aroused* was avoided as it might be used to refer more or less to the entire horizontal axis of the circumplex model.

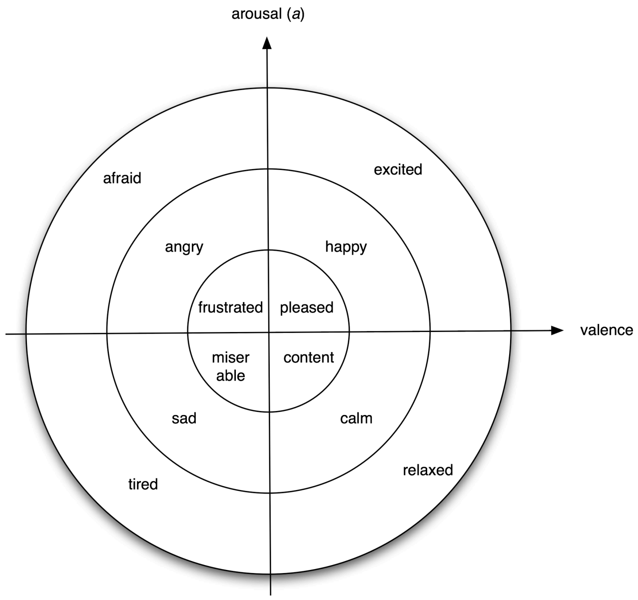

Figure 5. Quadrants with 12 discrete affective adjectives mapped from the circumplex model (*afraid*, *angry*, *frustrated*, *excited*, *happy*, *pleased*, *tired*, *sad*, *miserable*, *relaxed*, *calm*, and *content*).

The 'main' emotions (*sad*, *calm*, *angry*, and *happy*) can all be seen in the second level (*v1, a2*), (*v2, a2*), (*v1, a5*), and (*v2, a5*) of Figure 5. Thus, an emotional trajectory moving from *pleased*, via *happy*, to *excited*, is represented by a vector which gradually increases in arousal whilst maintaining positive valence: (*v2, a4*), (*v2, a5*), (*v2, a6*). This trajectory, and the underlying co-ordinates, can then be dynamically derived from gameplay in real-time and used to generate music in response to the intended emotional trajectory via a series of musical parameter constraints, as shown in Table 1.

Table 1 shows the corresponding musical feature mapping used to create a range of 12 affective targets from 6 discrete musical features (tempo, rhythmic density, modality, articulation, mean pitch range, and mean spectral features or more loosely, timbre). The precise ranges used in each feature can be determined by user interaction with a graphical interface (setting appropriate duration ranges for articulation, and determining appropriate tempo, pitch, and spectral outliers), or these values can be informed by the input ranges present in the MIDI representations used as source material.

Cartesian co-ordinate	Example descriptor	Musical parameter mapping
1. (a1, v1)	*Tired*	Low density, slow tempo, minor mode, legato articulation, low mean pitch range, low spectral range (dark timbre)
2. (a2, v1)	*Sad*	Medium density, slow tempo, minor mode, legato articulation, medium mean pitch range, low spectral range (dark timbre)
3. (a3,v1)	*Miserable*	Medium density, medium tempo, minor mode, legato articulation, medium mean pitch range, medium spectral range (clear timbre)
4. (a4, v1)	*Frustrated*	Medium density, medium tempo, minor mode, staccato articulation, medium mean pitch range, high spectral range (bright timbre)
5. (a5, v1)	*Angry*	High density, fast tempo, minor mode, staccato articulation, high mean pitch range, high spectral range (bright timbre)
6. (a6, v1)	*Afraid*	Medium density, medium, minor mode, staccato articulation, high mean pitch range, high spectral range (bright timbre)
7. (a1, v2)	*Relaxed*	Medium density, slow tempo, major mode, legato articulation, low mean pitch range, medium spectral range (clear timbre)
8. (a2, v2)	*Calm*	Medium density, slow tempo, major mode, legato articulation, medium mean pitch range, medium spectral range (clear timbre)
9. (a3, v2)	*Content*	Medium density, medium tempo, major mode, legato articulation, medium mean pitch range, medium spectral range (clear timbre)
10. (a4, v2)	*Pleased*	Medium density, medium tempo, major mode, staccato articulation, medium mean pitch range, medium spectral range (clear timbre)
11. (a5, v2)	*Happy*	High density, medium tempo, major mode, staccato articulation, medium mean pitch range, high spectral range (bright timbre)

Table 1. Showing cartesian co-ordinates, suggested affective descriptors, and the corresponding musical parameter mappings used by the generative model

4 DEMONSTRATION

A short excerpt of monophonic music generated by the pilot system is shown in Figures 6 and 7 (for readers who are not familiar with standard notation, audio examples of the system can be heard at cmr.soc.plymouth.ac.uk/bcmi)

Figure 6. Excerpt of seed material, condensed to a monophonic piano arrangement, taken from *Brandenburg Concerto No. 5*, J.S. Bach.

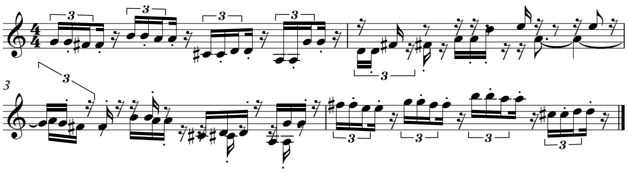

Figure 7. 'Lower density' excerpt created by Markov permutation of measures from seed edit, with lower density index used as the basis for selection of rhythm trees. Note the use of triplets to force a pattern from the latter half of the first line, Figure 6, into the output rhythm tree.

The generated output contains many of the thematic musical elements of the source material but is nevertheless unable to adequately represent the overall musical structure. Structural issues can be somewhat addressed in this system by using specific seed passages which correspond to narrative sections, for example using sources for the beginning, middle, and end of the narrative as in the basic branching example above. The compositional complexity involved in maintaining novelty in the soundtrack can thus be mediated by the continuous and real-time nature of the generation, whilst particular affective trajectories can be followed by manipulating the specific features determined in the affective mapping. This system could now be the subject of a rigorous perceptual evaluation such that the specific affective mappings used can be evaluated and perhaps adjusted. It is likely that the affective space represented in Figure 5 is not linear, and a game-specific evaluation would be necessary in order to determine any multimodal affect created by the act of player interactions as well as being presented with combined visual, auditory and tactile stimuli (such as in the case of vibrating controllers etc).

5 CONCLUSIONS

A pilot system for the dynamic generation of affectively-driven adaptive soundtrack creation, in response to a continuously varying real-time emotional trajectory in 2D is proposed, specifically to address the problems of real-time soundtrack creation faced by composers developing music for games. The system has potential to tackle some of the problems faced by traditional composition approaches, and offers a higher affective resolution than that which can be offered by branching systems (without a high degree of compositional complexity and accompanying computational load). Solutions which offer pixelation of source material to a high degree are also computationally expensive and compositionally complex to work with. Real-time generation offers the possibility of adaptive music which may help to sustain player immersion by maintaining novelty.

The pilot system can be used to generate near-infinite amounts of music without looping and with nonlinear emotional trajectories across 12 cartesian affective co-ordinates. In the future, by combining the system with biosensors and neurofeedback (for example, electroencephalogram (EEG) as illustrated in [20]), the system could be adapted to respond individually to the perceived affective state of the player. By training the system with short excerpts of material corresponding to decision points, abrupt transitions and codas can also be avoided. However, seed materials with branching structures are not well represented by a state transition matrix such as the one used in the Markov chain analysis presented here. In an affective perceptual evaluation of the pilot system, the issue of overall musical structure in this system is of paramount importance when considering future work in this area. Therefore a means for generating structural descriptors from a hierarchical analysis of the source material, before categorising sub-structures and classifying them by means of their affective correlations for use in the subsequent generation is suggested as one possible avenue for further work, such that the emotional trajectory of the gamer is used as the central musical decision-making process.

6 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of EPSRC grants EP/J003077/1 and EP/J002135/1.

REFERENCES

[1] S. D. Lipscomb and S. M. Zehnder, "Immersion in the virtual environment: The effect of a musical score on the video gaming experience". *J. Physiol. Anthropol. Appl. Human Sci.*, vol. 23, no. 6, pp. 337–343, November 2004.

[2] M. Chion and C. Gorbman, *Audio-vision : sound on screen*. New York, NY : Columbia University Press, 1994.

[3] K. Collins, "An Introduction to the Participatory and Nonlinear Aspects of Video Games Audio". *Essays on Sound and Vision*, Eds. Stan Hawkins and John Richardson. Helsinki: Helsinki University Press. pp. 263-298, 2007.

[4] A. Berndt and K. Hartmann, "Strategies for Narrative and Adaptive Game Scoring" in *Proceedings of the Audio Mostly Conference, Rontgenbau, Germany*. September 2007.

[5] M. Kamp, "Musical Ecologies in Video Games" *Philos. Technol.*, vol. 27, no. 2, pp. 1–15, June 2013.

[6] B. L. Jacob, "Algorithmic composition as a model of creativity" *Organised Sound*, vol. 1, no. 3, pp. 157–165, December 1996.

[7] G. Nierhaus, *Algorithmic composition paradigms of automated music generation*. Wien; New York: Springer, 2009.

[8] N. Collins, "Musical Form and Algorithmic Composition" *Contemp. Music Rev.*, vol. 28, no. 1, pp. 103–114, February 2009.

[9] I. Wallis, T. Ingalls, and E. Campana, "Computer-Generating emotional music: The design of an affective music algorithm" In *Proceedings of the 11th International Conference on Digital Audio Effects, Espoo Finland*, September 2008.

[10] J. Rubisch, M. Jaksche, and H. Raffaseder, "Generative Music for Media Applications (GeMMA)–Towards Automated Design and Production of Media Related Music" In *Proceedings of the Audio Mostly Conference, Glasgow, UK*, September 2009.

[11] J. A. Russell and L. F. Barrett, "Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant" *J. Pers. Soc. Psychol.*, vol. 76, no. 5, p. 805, May 1999.

[12] J. A. Russell, "A circumplex model of affect" *J. Pers. Soc. Psychol.*, vol. 39, no. 6, p. 1161, December 1980.

[13] K. Hevner, "Experimental studies of the elements of expression in music" *Am. J. Psychol.*, vol. 48, no. 2, pp. 246–268, 1936.

[14] E. Schubert, "Measuring Emotion Continuously: Validity and Reliability of the Two-Dimensional Emotion-Space" *Aust. J. Psychol.*, vol. 51, no. 3, pp. 154–165, December 1999.

[15] K. R. Scherer, "Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them?" *J. New Music Res.*, vol. 33, no. 3, pp. 239–251, September 2004.

[16] A. Lamont and T. Eerola, "Music and emotion: Themes and development" *Music. Sci.*, vol. 15, no. 2, pp. 139–145, July 2011.

[17] D. Williams, A. Kirke, E. R. Miranda, I. Daly, E. B. Roesch, and S. J. Nasuto, "Investigating Affect in Algorithmic Composition Systems" *Psychol. of Music,* vol. 42, no. 6, November 2014.

[18] D. Williams, A. Kirke, E. Miranda, I. Daly, J. Hallowell, F. Hwang, E. Roesch, and S. J. Nasuto, "Evaluating perceptual separation in a pilot system for affective composition" in *Proceedings of the 40th International Computer Music Conference*, Athens, Greece, September 2014.

[19] C. Ames, "The Markov process as a compositional model: a survey and tutorial," *Leonardo*, vol.2, no. 22, pp. 175–187, February 1989.

[20] I. Daly, A. Malik, F. Hwang, E. Roesch, J. Weaver, A. Kirke, D. Williams, E. Miranda, S. J. Nasuto, "Neural correlates of emotional responses to music: An EEG study" *Neurosc. Letters,* vol. 573, pp. 52-57, June 2014.