Soundtrack Localisation: Culturally Adaptive Music Content for Computer Games[15]

Ian R. O'Keeffe

Localisation Research Centre, Centre for Next Generation
University of Limerick, Co. Limerick, Ireland.
ian.okeeffe@ul.ie

Abstract

This paper focuses on the localisation and adaptation of one particular aspect of the computer game: the soundtrack. Sometimes bespoke, sometimes selected from commercial releases, it provides a background, supporting role in creating atmosphere and supporting the emotive state of the game space. But how often is the target market considered when selecting appropriate musical content? Is it possible to use this almost subliminal channel into the game player's consciousness to increase his awareness of what is happening around him, and give him a feel for his character's emotional and physical wellbeing? This paper presents a novel approach for transforming the soundtrack - via a system the author originally created for the purposes of capturing and recreating emotive content in music.

Keywords: Soundtrack, Computer Games, Emotive Musicology

[15] This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie) in the Localisation Research Centre in the University of Limerick.

Introduction

Computer games are, in many ways, the ultimate multimedia experience. For the purpose of this paper I will consider computer games to be "…any forms of computer-based entertainment software, either textual or image-based, using any electronic platform such as personal computers or consoles and involving one or multiple players in a physical or networked environment." (Frasca, 2001). Games exist across many genres: Action and Adventure, Driving and Racing, First-Person Shooter (FPS), Platform and Puzzle, Role Playing Games (RPG), Strategy and Simulation & Sports and Beat -'em-ups being one available list (Berens, & Howard, 2001). The visuals of a computer game are non-linear, reacting to the interactions of the player, as are the sound effects and any haptic feedback generated by the control devices (joysticks, button controllers, motion-sensitive wireless gesture sensing devices and so on). The sonic landscape is generally in stereo at a minimum, and often in full surround, to create the illusion of being within the space depicted on-screen and it is made up of many layers such as incidental sound effects, dialogue, status warning sounds and the soundtrack. "Sound is more immersive than graphics. While graphics will draw you in to a scene, the sound going on in the background will create a reality in the player's mind that can never be done with graphics alone." (Howland, 1998). Computer games have more in common with cinematography in this respect than most other computer applications, as sound is regarded as such a central aspect of the user experience. Indeed, in most other computer use scenarios sound is not perceived as being that important and may even be seen as an issue. Consider the laptop user on a crowded train who navigates to a website featuring loud music and has to quickly mute the speakers for fear of offending fellow travellers. In contrast an enthusiastic gamer will have a good set of headphones, or even a dedicated home surround sound speaker setup, to help enhance the gaming experience. This is because the principle of the suspension of disbelief, a concept first introduced by the poet Samuel Taylor Coleridge, is considered to be "one of the fundamental tools in creating a successful game design" (Crosignani, Ballista, & Minazzi, 2008) and the more immersive the experience the more likely the game player is to become enmeshed in the narrative, the characters and the virtual setting. This is why it is so important to avoid shattering this spell - through factual inaccuracies, inappropriate dialogue, unrealistic game-play or incorrect soundtrack elements - as once lost it is very difficult to regain the focus of the player. To further discuss the requirements for creating a computer game I shall now move on to investigate the game elements that make it up in more detail.

Design

From a design perspective one of the first elements to be considered is the overall narrative along with subordinate elements such as game play features, character information and so on. When it comes to the actual physical interaction with the game there are three basic groups: graphics, controls and haptic feedback, and sound, or more basically: things you can see, things you can feel, and things you can hear. Graphics can then be further broken up into game area graphics, status displays and menus. Game area graphics present the virtual world the player inhabits while playing the game. A recent development is the increasing availability of 3D graphics, necessitating the use of 3D glasses and special screen hardware. The status displays present information to the player such as energy left, location on a map, weapons available, time, game level and so on. These can often be customised in terms of positioning, content and visibility. Menus are normally used for navigating selection choices, something not normally done during intensive interactive game-play. Controls and haptic feedback make up the physical interface between the player and the game, the game control interface. Examples include the keypad used for controlling characters, specialist input devices like car steering wheels, pedals and gear levers, wireless gesture capture devices (Nintendo Wii), and more recently, human gestures captured by image analysis via a camera mounted on the games console (Microsoft Kinect). Haptic feedback from these devices can be in the form of vibrations, or force feedback in the form of resistance or 'kick-back' such as would be experienced through a steering wheel in a car. The final group of elements, sound, are outputted via the headphones or the speaker system. These consist of a number of layers: effects, dialogue, and the soundtrack. The effects typically relate to players actions, and events that occur within the game play. Dialogue is a special case relating to actual speech made by characters within the game. Both of these layers are positioned within the game space, both logically and spatially. In other words, the nature of the acoustic space in which the player finds himself will generally affect these sounds in terms of echo or reverberation, and stereo or surround positioning. The soundtrack is the final layer, and does not share this close-coupling with the game space. It exists to set the mood and enhance the game experience. A soundtrack is generally regarded as effective if it manages to remain 'unnoticed' by the player/audience, remaining at a subliminal level, and reinforces the on-screen action rather than distracts from it. However, like many other aspects of the computer game, it still requires modification and fine tuning to be acceptable to different locales and cultures.

Localisation

Computer games, like all computer software and related digital content, face localisation requirements if they are to be commercially viable in the global marketplace. In some cases these requirements can be met just by translating any text visible to the player, such as menus, signage within the game itself, or the status display, but more complex games will normally require a higher degree of modification, such as re-recording any dialogue in the required target languages. There is also the issue of cultural modification, one example being catering for cultural gaming conventions such as the preference in Asian game characters for more child-like characteristics (anime or manga) versus the more adult characters in Western games (Lara Croft) (Trainor, 2003). Another related cultural issue is the degree of freedom in depicting sexuality or violence. Europe typically shows much more concern for violence in games than it does for nudity, with the opposite being true of the US market. Germany for example has very strict laws relating to the depiction of violence in computer games, in particular where blood and gore are concerned (Chandler, 2005; Dietz, 2006). The soundtrack itself may also face issues, either with regard to broadcasting rights for a geographical area for a particular piece of commercial music or perhaps from a cultural perspective where the style of music may not suit the target audience. Another area of cultural concern, and the primary focus of this paper, relates to the emotional reaction of a player to the soundtrack if it is intended to communicate emotive or expressive content relating to the game play itself via the soundtrack. As music cognition and emotional response is a complex area in its own right, it is necessary to first present an overview of the process of creating a soundtrack for a computer game.

Soundtrack creation

The soundtrack's task is to help create atmosphere and enhance the gaming experience, either by supporting the visual scenes presented or by giving feedback on the performance of the player. The musical content for a game soundtrack can be selected from current commercial tracks, but in a lot of cases it is composed specifically for the game being created as this allows it more flexibility in dealing with specific requirements, such as the non-linear nature of game play when compared to the controlled environment of a movie/film. It is in this area of soundtrack composition that I will place my focus here. It is a useful 'side-effect' that bespoke compositions also have the added benefit of avoiding expensive royalty payments where global sales are concerned.

I will assume the main musical themes have already been composed at a conceptual level, to suit the marketing and style requirements of the particular genre of game, and the focus is now on tailoring the soundtrack to fit in with the game play itself. It must be stressed, however, that while some aspects of the soundtrack can be composed in advance many situations require it to be altered or amended in a non-linear manner. This makes such music well suited to a procedural composition approach. It should be noted that the focus here is not on the interactive elements (Collins, 2009), sound effects events triggered directly by the player through their actions via their input device such as the swishing of a sword or the sound of footsteps. Instead, it is the adaptive audio that is under investigation. These adaptive audio events are triggered by a game's sound engine based on certain in-game parameters such as general locations, time of day, 'camera' angle, or player properties like health or skills, and as such have a much more indirect link to the actions of the player. They use a core of musical content to generate the required duration of soundtrack in real time to fit in with the specific requirements of each player as they progress through the game. Another contrast with the linear, fixed movie soundtrack is the amount of music that needs to be composed due to the amount of time it takes to complete most modern games, and of course online multi-player games may never actually end at all. One good example of a dynamic composition engine is the iMUSE system (Interactive Music) developed and used by LucasArts (Land, & McConnell, 1994) in such titles as Totally Games / LucasArts' "X-Wing" series. The abstract from the patent describes the technology as follows: "A computer entertainment system is disclosed for dynamically composing a music sound tract (sic) in response to dynamic and unpredictable actions and events initiated by a directing system in a way that is aesthetically appropriate and natural". Then an example is given to differentiate the technology from existing approaches, where a fight scene features three music sequences: fight music (looped), victory music and defeat music: "The fight music, rather than playing along unresponsively, can be made to change mood of the game in response to the specific events of the fight. For example, certain instrument parts can signify doing well (for example, a trumpet fanfare when a punch has been successfully landed), while others can signify doing poorly". It is also possible to transpose the music in pitch as the fight reaches its climax, and to incorporate transitional phrases to avoid an abrupt jump from the fight loop to the victory music. Some more examples of the soundtrack reacting to game stimulus: the game 'Super Mario Bros.' (Nintendo), where the actual tempo of the music increases as time runs out for the player, or the practice within games to gradually reduce the intensity of the music over time or maybe stop it altogether to avoid irritation due to repetition if a player has become stuck in a particular location for a long time.

So it can be seen that the soundtrack is a key aspect of any computer game, and can react to game trigger events to allow it to be modified in real time to convey more context information. It is also able to adapt to fill indeterminate time slots, and reflect aspects of the player's status and the general surroundings of the game location. But it could do so much more, given the power music has to influence our emotions and our ability to track multiple information streams in parallel via our hearing as evidenced by the "cocktail party effect" (Arons, 1992). The fact that the soundtrack data is accessible from a computational standpoint means that it becomes a prime candidate for emotive adaptation. What will be presented here is an approach for capturing and re-using emotional templates for music, thus providing the computer game industry with a mechanism for altering soundtrack content to project any mood or emotion required by the game developers or indeed by the players themselves. A search of some online gaming forums identified some interesting threads, one example being from www.neogaf.com that requested members to list video game music that made them happy.

While the names of the games and songs listed by members were of genuine interest (Brawl - Yoshi's Story Ending Theme, Tekken 2 - Michelle's Theme, Marvel Super Heroes vs Street Fighter - Sakura's Theme, Final Fantasy V - Dragon Flight theme etc.), it is forum interactions such as that quoted below that really highlight the need for care in soundtrack selection.

> Forum member #1:
> "Also this one makes me happy everytime i hear it Final Fantasy VIII - Balamb Garden"
> Forum member #2:
> "That song always makes me sad and nostalgic. -sniffle-"

From a localisation perspective, this raises the question of universality in emotion in music. Do we need to localise or culturally adapt the soundtrack to suit each locale, culture or demographic?

In terms of positioning this research is best viewed as being placed at the intersection of musicology, music technology and psychology. More specifically, this research is placed within a sub discipline of musicology, that of music cognition. This area is concerned with the study of music as information from the viewpoint of cognitive science. This discipline shares the interdisciplinary nature of other fields such as cognitive linguistics. Music technology is the result of applying computers and other forms of technology to the creation and adaptation of music. It should be noted that the initial focus of the research presented here was on the capture of emotional content in music for one particular culture, listeners to western art music, but the data gathered in this particular application demonstrates a lot of promise for the expansion of this technical approach into other areas of music analysis and modification.

Proposal

The proposal is to put a mechanism in place that facilitates the modification of the existing soundtrack to include status information from the game such as strength and threat level as well as information relating to the player's emotive state, so that the player becomes aware, almost subliminally, of what is happening around him and has a feel for his character's emotional and physical wellbeing. In essence, I would see a situation where the soundtrack accompanying a computer game responds actively to various environmental variables and therefore enhances the game experience. This layer of modification would be placed on top of any existing music engine, and would act upon the original soundtrack. For example, if the player is acting particularly aggressively then anger could be introduced to the soundtrack. Low energy levels could also be modelled by capturing templates from users asked to create modified music that in their view represents tiredness or fatigue. Completeness of each level of a game could also be modelled, perhaps by increased tempo or intensity (by altering the stresses applied to individual notes, shortening the sounding times and so on). There is no reason to view the process as one directional either, with the player's actions only influencing the soundtrack. The opposite could also be argued to be true as research has shown that music has the ability to induce mood in a listener (Pignatiello, Camp, & Rasar, 1986). If there is an inherent threat to the player which may not be visible then subtle hints of menace, danger or fear could be introduced to the music, the amount of influence depending on proximity or perceived threat level, and the player would therefore feel the effect of this emotive content himself by listening to the music, thus increasing the immersive nature of the game experience. Moving elements of status information from the screen to the soundtrack would also have the benefit of freeing up screen real estate, and thus also helping to enhance the suspension of disbelief by removing text from the view the player sees.

Some suggestions for possible mappings from status information to musically mapped emotions could be: High energy levels - Strong/Happy, Finished a level - Happy, Low energy - Sad, Running away from enemy - Fearful, Aggressive game play - Angry, Low on ammunition - Worried and so on. It is also plausible to consider mapping emotions from the facial expressions of hero/player if so expressed graphically, as long as there was access to such status via some tag or variable within the game's code base.

Localisation Impact

From a localisation perspective the question that must be asked is: how much thought goes into the cultural suitability of such musical content? The emotive template needs to be representative of the player's locale so that it can be correctly recognised and processed as it is quite possible that different cultures may derive differing meanings and understanding from music due to their cultural conditioning, and an acceptance that the structures of Western Art Music can be viewed as universal is a dangerous assumption to make (O'Keeffe, 2009). This implies a requirement to ensure the correct set of emotional templates is selected depending on the player's demographic. As an example, perhaps music deemed happy in some cultures may be perceived as melancholy in others. It is therefore necessary to review the existing research in musical cognition, particularly any research that focuses on cross-cultural differences in comprehension and understanding, prior to describing possible mechanisms for creating the emotional modification of music, because if there is no evidence of such cultural diversity then the concept of 'music localisation', for this is effectively what we would be considering, becomes superfluous.

The majority of the work in this area tends to focus on the psychological aspects of musical understanding, and the physiological results, rather than on how the music itself may be adapted to suit differing cultures or locales. The research also generally takes the form of passive studies of test subjects and their reactions to pre-prepared musical data, rather than an active approach where the test subjects are able to change the music themselves. Looking firstly at psychological approaches, Gregory and Varney (1996) conducted a study that looked at the affective response of subjects to music from different cultures and found that listeners “brought up in the Indian cultural tradition have difficulty in appreciating the emotional connotations of western music”. Walker (1996) states that “understanding the music of another culture requires assimilation of the influences affecting musical behaviour as much as of the resultant musical products”, suggesting that cultural conditioning plays a part in how a listener understands music. Different cultures and ethnic backgrounds also show preferences for different types of music for musical therapy, as demonstrated in a study of the music selected by medical patients to help with post-operative pain relief (Good, Picot, Salem, Chin, Picot, & Lane, 2000). Moving on to neuro-science, a study (Morrison, Demorest, Aylward, Cramer, & Maravilla, 2003) of human brain activity captured using functional magnetic resonance imaging (fMRI) showed that there were activation differences between Western (familiar to test group) and Chinese (unfamiliar) music based on training. Trained listeners showed extra activation “in the right and left midfrontal regions for Western music and Chinese music, respectively”. It would therefore seem that we react differently to unfamiliar musical styles or traditions whether we want to or not. In line with this review of the literature, the author has set up an online survey (www.localisation.ie/resources/MusicLOCweb/) to answer the question: Do different cultures hear music differently, and is the emotional content in music universal, or dependant on locale? The experiment requires the test subject to listen to a simple piece of music, and then to modify its mood using basic controls for speed, rhythm, dynamics (how loud it is played), note length (short or long) and scale (major or minor). The aim is to end up with four versions of the piece that the test subject judges to sound Happy, Sad, Angry or Fearful. The survey is still active, as data gathered to date has been mostly from the US/European locales, so a more diverse dataset needs to be collected before any true trends can be identified. The data does demonstrate broad agreement on emotive templates by region, however, which is encouraging. What is reassuring in the research discussed here is that there does seem to be evidence that music is not the universal language that it is often thought to be and that there is a place for musical localisation within game soundtracks, particularly when combined with emotive profiling.

When the scope of game play is expanded to include Massively Multiplayer Online Games (MMOGs) then the opportunities to create adaptive soundtracks expand in parallel allowing the mirroring of player status, such as health or mood, in the soundtrack they broadcast to other gamers as well as the music they hear themselves. This could be achieved by giving each player their own distinctive soundtrack carrying subliminal information about their mental state, something along the lines of the leitmotif used by Wagner, and more recently by John Williams in his Star Wars soundtracks, where one character thinks of another character or of an emotion and the soundtrack hints at this. This would allow each player to not only gain information about their own status from their soundtrack but also to pick up on the presence, and emotional make-up, of other game players in their proximity. Of course, this also increases the possibility for confusion if the cultural background of each player is not taken into consideration when presenting them with emotive musical information. It becomes apparent that some form of adaptation would be necessary where the soundtrack is concerned to ensure that the perception of each individual player matches that planned by the games developer so individual templates, like personal profiles, could be held by each player and used to present them with their own culture-centric mood-corrected version of the overall game soundtrack. Perhaps consider the analogy of having all textual visual signage presented in your own language as you look at it, even though another player standing 'beside you' in the virtual space may well see the same text in a different language. Of course, such conceptual discussion requires a physical mechanism for modifying a music stream if it is to be moved from the realms of conjecture to reality, and the next section presents just such a framework.

Mechanism

The objective for the research was to create a system that would allow any listener to directly create emotion in music. This was in response to a review of existing emotional research methods for capturing emotive data where it was found that they varied from asking test subjects to listen to live performances and report on them (Downey, 1897; Gilman, 1891) through to judging pieces of music composed to express the required emotions (Thompson, & Robitaille, 1992; Rigg, 1937) or phrases selected from existing classical compositions (Gundlach, 1935) and finally to asking professional musicians to play pieces in various emotions and study the reaction of the test subjects to these pieces (Gabrielsson, & Juslin, 1996). What was interesting was that very few of these experiments required that the test subjects create the emotional content themselves, instead asking them to play the role of passive listener, and if musical input was required the user was normally required to be a skilled musician. The approach presented here strives to overcome these shortcomings by creating a series of low-level musical operations from scratch, bypassing the analysis stage, and creating a system that allows the user build emotive modification presets or 'templates' with these operations, thus synthesizing the emotion. These 'templates' of the user actions were then captured for analysis and potential re-use. The advantages of this include: no pre-requisites of musical ability, either as a performer or theorist, no dilution of emotive content through syntactical loss in a process of description or through loss of immediacy caused by reflection after listening (Meyer, 2001), and direct results. Allowing users to construct the conversion "templates" themselves also avoids the issue of direct influence and bias on the part of the researcher.

Transformations

The music transformations were as follows: Tempo - The speed of the performance. Pitch - This transformation performs transposition. Rhythm - The simplest rhythm for a piece is to have each note equal in length, creating a regular pulse. Ways to alter this pulse include staggering it, making it act against the natural rhythm of the piece in the form of syncopation, or having a random variance from note to note, creating unexpected results. Timbre - Which instrument to use to perform the piece. Harmony - Reinforce the melody with an accompanying voice positioned at a harmonic interval above the primary melody. Accompaniment - Separate from the harmonic accompaniment mentioned above, this option gives the user the ability to add an accompanying bass line. Dynamics - How loud the piece is played. Drum Rhythm - Does the addition of a drum rhythm affect the emotive content of a piece, and if so, do different rhythms have different effects? Attack - Individual stressing of beats in a bar: the inhuman precision and evenness of a machine, the carefully stressed beats of a skilled musician, or simple random stressing. Articulation - From staccato, for very short stabbing notes with significant gaps prior to the next note, through to legato, where each note begins to link up with the next note in the sequence. Scale - Major, Minor, Pentatonic, Blues & Gypsy scales.

Emotions

The list of basic emotions for the study was compiled after analysing the proposals of several psychologists: Clynes (1980), Izard (1991), Plutchik (1980; 2001), Scherer (1995), Schopenhauer (Gale, 1888), and Shand (1914). Of particular interest was Plutchik's classification system, with eight sectors representing eight primary emotion dimensions arranged as four pairs of opposites. For this reason Robert Plutchik's model was selected over those of Shand, Izard and Clynes as it includes a pairing of opposing emotions, thus allowing further comparisons to be made between emotions as well as analyzing each emotion separately. The list used:

> Joy, Sadness, Anger, Fear, Acceptance, Disgust, Surprise, Anticipation and No Emotion (the control emotion).

Notation

The system described in this paper uses the MIDI music file format for storage and manipulation of the musical content. It was selected because of its wide availability, portability and small storage footprint, and because it stores music in a symbolic way. It does have some limitations, and perhaps other notation methods would need to be considered in the future, but MIDI is certainly a good place to start given its wide acceptance worldwide and its use within the computer gaming industry itself.

Studies & Results

The Main Study (see (O'Keeffe, 2009) for a more in-depth explanation) consisted of nine participants. A Cooperative Evaluation (Monk, Wright, Haber, & Davenport, 1993) approach was taken, in which each user was encouraged to see himself as a collaborator in the evaluation and also to actively criticise the system rather than simply suffer it, thus allowing the evaluator (myself) to clarify points of confusion so maximising the effectiveness of the approach. Cooperative Evaluation is a variant of Think Aloud (Jorgenssen, 1989), in which the user performs a number of tasks and is asked to 'think aloud' to explain what they are doing at each stage and why. The user's actions were recorded via paper-based notes and via the command tracking script system built into the main system for each of the tasks. A task list was prepared and given to each of the nine participants prior to the start of the study, and consisted of a walkthrough of the system followed by the creation of each of the nine required emotions. For the purpose of isolating any significant findings the Chi Square Goodness of Fit test was applied to the data gathered, given its nominal/categorical nature (Figure 1). An alpha value of 0.05% was selected for the test for significance, although the sample size may be considered to be a little on the small size for robust findings. Even so, it gives a good indication of the trends in the data gathered. The results of this study show that the proposal of collecting emotive data from test subjects using low-level musical transformations definitely has merit, and some interesting trends are apparent (Tables 1 through 11).

$$\chi^2_c = \sum \frac{(O_i - E_i)^2}{E_i}$$

Figure 1. Chi Square Goodness of Fit formula.

Tempo is definitely a decisive factor, particularly in Joy, Sadness, Anger and Surprise. Some more examples of emotional opposites in the data, as suggested by Plutchik in his arrangement of the primary emotions as pairs of opposites, can be seen in Pitch between Joy and Sadness, and Anger and Fear; in Rhythm between Joy and Sadness, Acceptance and Disgust, and Surprise and Anticipation; in Dynamics for all emotional pairs; and also in many aspects of Attack Length and Articulation. Scale shows an almost bipolar split between Joy and Sadness, and also shows a lot of variance across the other emotions. Moving the focus away from each transformation and onto individual emotional templates, when they are compared in the combined chart (Figure 2) the variance between them is apparent. This would suggest that each emotion is mapping to its own set of preferences, and also suggests that there is validity in attempting to extract emotional content from music in this manner.

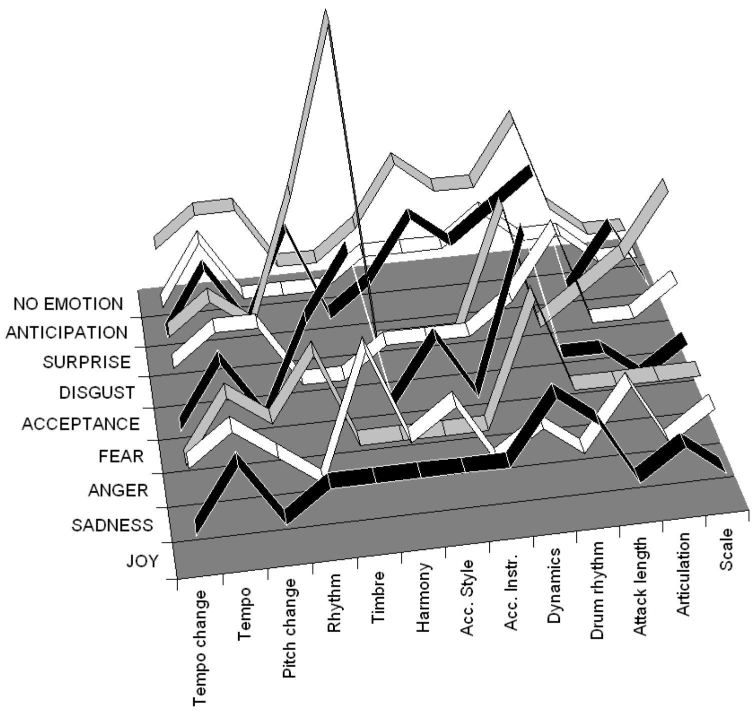

Figure 2. A combined chart of individual emotional templates.

The templates produced by the main study presented an excellent opportunity for verifying the data gathered by the system simply by running the experiment in reverse. To this end, a study was created using a short piece altered emotionally by the system using averaged emotional templates gathered from the main study, and the test subjects were asked to categorise these pieces. The study involved listening to the nine pieces - each representing a different emotion - and categorising them by Emotion. The results were split into two groups, those who were already familiar with the system, and those who had never seen it and were judging the pieces provided purely on emotion. The results are displayed here in the form of confusion matrices (Tables 12 & 13).

The results (Chi Square analysis with an alpha value of 0.05% as before) showed a significant outcome for the data collected for each emotion, although not all emotions were correctly identified. What is also interesting is how closely the data matches between the two groups, suggesting that there is indeed a recognisable set of emotions in music whether you are familiar with the system and its processes or just a listener to any emotionally altered piece.

Discussion

The results gathered by this research show much promise for an interactive approach to music analysis and alteration. The benefits are observable: clear, empirical data captured from precise on-screen user decisions, avoidance of confusion through bias, introspection, and descriptive issues, ease of usability across a broad musical ability spectrum and easy capture of emotive templates for re-use. The follow-up study also demonstrated that listeners were able to correctly identify the intended emotion or mood in pieces of music that had been altered by the templates gathered by the system.

To summarise the capabilities the approach provides:

> The ability to capture a user's preferences in the form of a template in response to a request to induce an arbitrary mood, emotion or categorization in any piece of music;
>
> The ability to store this template for re-use on other musical material as required;
>
> The ability to generalize the data gathered across a number of users for any specific template descriptor: for example, sadness;
>
> This then implies the ability to segment any generalized findings into any required demographic, such as locale, culture, game genre, gender and so on.

Now that the concept and practice of capturing emotive templates for music has been demonstrated, the next phase is to ask how this system could be leveraged in a computer gaming environment as was proposed earlier. First of all, let us consider how the system currently operates and compare this with the requirements of a computer game soundtrack generation engine. In the system presented here the target of the emotive modification process is a static MIDI file and the transformations are applied to the entire piece in one pass. To be able to react in real time the processing would need to be re-targeted to alter the music as it is being produced by the music engine, perhaps being a part of the workflow for that process. Any transformation functions that rely on knowing how far into the piece they are would probably have to be either removed or substantially remodelled. It would also be possible to implement modification functions that act on a bar of music at a time and then alter the music at the start of the next bar in the soundtrack. As the emotive information is not as immediate as a sound effect for, say, a sword swish then this approach could be quite effective. Another issue is the use of MIDI itself, but the transformations are described logically in terms of musical alterations so could be applied to any symbolic musical representation method.

Moving on from the technical considerations of system integration, part of the preparation for emotive soundtrack adaptation in the creation of a new game would involve establishing what profiles are required in terms of moods or emotions. A set of categories for the templates required for any given game scenario would need to be proposed, such as: Basic Emotions (Happy, Sad, Angry, Fearful and so on), or game-play variables (Fatigue, Adrenaline, Amount of game level complete, Weapons gathered and so on). There would also be a requirement for sub-categories within each culture or locale, as the templates for given emotions could quite possibly differ across cultures, and this must be taken into account as preparation for producing localised versions of games. These categories could always be expanded later as required. Decisions would need to be made as to what game data should be tracked, and how it should be mapped onto the soundtrack. For example, consider progression through a stage or level in a game, performance of a player, energy levels, or behaviour analysis. At a higher level, the interaction of differing basic emotions or signals could have interesting results, such as the combination of anger and disgust leading to more complex emotions such as contempt. This could also create highly individual and personal soundtrack content if the modification process was made up of a product of a series of individual templates, each weighted dependant on influence, strength, distance or whatever.

Possibly the biggest area of work would relate to the creation of the templates themselves in all the required categories, sub-categories and cultures. The game authors would be responsible for producing the definitive version for each culture/locale, but there would also be the possibility of user-created profiles, where individual players could save their own favourite modification profiles. A further development along this path could see players being able to upload their profiles, or share them with fellow gamers. This could be regarded as being a variant of the crowd-sourcing model, where users would be able to create their own cultural templates for the soundtrack of the game, possibly with the option of uploading these templates for use by the global community.

The term 'Crowdsourcing' was first coined by Jeff Howe (2006) and later defined as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call." (Howe, 2009). Howe also refers to it as "the future of corporate R&D", citing the example of InnoCentive, the "research world's version of iStockPhoto" (Howe, 2006). In contrast, what is of more interest here is crowdsourcing motivated simply through personal desire to make a contribution, such as demonstrated by contributors to the translation of Facebook. Their reward is recognition from their peers, and perhaps personal satisfaction. No money changes hands.

What is proposed here is the incorporation of a simplified version of the music modification application embedded in the game itself that would allow any player to tweak the musical transformations that are connected to the game variables, and thus create templates for their own use. As an example, a player may see a latent threat in the game-play as being embodied by a minor key shift and a faster tempo. A further development would be allowing the sharing of these user profiles. The most extreme reading of this trend would be the creation of a website for a game to allow the management and collation of these profiles, thus opening up the possibility of facilitating the 'localisation' of the game soundtrack to match a player's own particular locale or culture, particularly if they feel the existing templates for emotions, moods or categories do not accurately reflect their preferences. For example, the default representation of anger in a wargame soundtrack may be completely at odds with the idea of musical anger in some cultures. In fact, it would be hoped that this form of 'national pride' would provide the motivation for contribution, as it has done for some minority languages on Facebook. Quality enforcement, and the avoidance of online 'vandalism', could be realised by including a peer voting system similar to that implemented by Threadless, the web-based t-shirt company (Brabham, 2008). This would allow visitors to the site to vote on the accuracy and appropriateness of existing templates. The data gathered could be of great value to the game authors, giving them a window into the requirements of the gaming community and helping with their localisation requirements for creating global versions of their software.

Conclusion

In conclusion, this research demonstrates the plausibility of creating a system to capture emotion-specific templates in music by involving the participant directly in the process of creating those templates. While there is still a lot of work to be done from a propagation standpoint this work presents a new research possibility in the fields of cognitive musicology, computer gaming and cultural adaptation - the possibility to perform emotive modification on music by culture or locale for the adaptation of game soundtracks - and the system constructed is a good foundation for the further development of such functionality.

References

Arons, B. (1992). A review of the cocktail party effect. *Journal of the American Voice I/O Society, 12*, 35-50.

Berens, K., & Howard, G. (2001). *The Rough Guide to Videogaming 2002.* London/New York: Rough Guides.

Brabham, D. (2008). Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence*, *14*(1), 75-90.

Chandler, H. (2005). *The Game Localization Handbook.* Massachusetts: Charles River Media.

Clynes, M. (1980). The communication of emotion: theory of sentics. In R. Plutchik, & H. Kellerman, *Theories of Emotion* (Vol. 1, pp. 171-216). New York: Academic Press.

Collins, K. (2009). An Introduction to Procedural Music in Video Games. *Contemporary Music Review*, *28*(1), 5-15.

Crosignani, S., Ballista, A., & Minazzi, F. (2008, October/November). Preserving the spell in games localization. *MultiLingual*, 38-41.

Dietz, F. (2006). Issues in localizing computer games. In K. Dunne, *Perspectives in Localization* (pp. 121-134). Amsterdam/Philadelphia: John Benjamins Publishing Company.

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

Frasca, G. (2001). Rethinking agency and immersion: video games as a means of consciousness-raising. SIGGRAPH 2001. Retrieved from http://siggraph.org/artdesign/gallery/S01/essays.html

Gabrielson, A. J. (1996). Emotional expression in music performance: Between the performers intention and the listeners experience. *Psychology of Music*, *24*, 68-91.

Gale, H. (1888). Schopenhauer's Metaphysics of Music. *New Englander and Yale Review, 48(CCXVIII)*, 362-368.

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

Good, M., Picot, B., Salem, S., Chin, C., & Picot, S. (2006). Cultural Differences in Music Chosen for Pain Relief. *Journal of Holistic Nursing*, *18*(3), 245-260.

Gregory, A. H., & Varney, N. (1996). Cross-Cultural comparisons in the affective response to music. *Psychology of Music*, *24*, 47-52.

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

Howe, J. (2006). The Rise of Crowdsourcing. Wired Magazine. *14*(6). Retrieved from http://www.wired.com/wired/archive/14.06/crowds_pr.html

Howe, J. (2009). *Crowdsourcing: A Definition.* Retrieved from http://crowdsourcing.typepad.com

Howland, G. (1998). *Game Design: The essence of computer games.* Retrieved from http://www.cpp-home.com/tutorials/

Izard, C. E. (1991). *The psychology of emotions.* New York: Plenum.

Jorgenssen, A. H. (1989). Using the think-aloud method in systems development. In G. Salvendy, & M. Smith, *Designing and Using Human-Computer Interfaces and Knowledge-Based Systems.* Amsterdam: Elsevier Science.

Land, M. Z., & McConnell, P. N. (1994). *Patent No. US Patent No. 5,315,057.*

Meyer, L. B. (2001). Music and Emotion: Distinctions and Uncertainties. In P. N. Juslin, & J. A. Sloboda, *Music and emotion: Theory and research* (pp. 341-360). New York: Oxford University Press.

Monk, A., Wright, P., Haber, J., & Davenport, L. (1993). *Improving Your Human-Computer Interface, A Practical Technique.* Prentice Hall International.

Morrison, S. J., Demorest, S. M., Aylward, E. H., Cramer, S. C., & Maravilla, K. R. (2003). fMRI investigation of cross-cultural music comprehension. *NeuroImage, 20*, 378-384.

O'Keeffe, I. R. (2009). Music Localisation: Active Music Content for Web Pages. *Localisation Focus - The International Journal of Localisation, 8*(1), 67-81.

Pignatiello, M., Camp, C. J., & Rasar, L. A. (1986). Musical mood induction: An alternative to the Velten technique. *Journal of Abnormal Psychology, 95*(3), 295-297.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik, & H. Kellerman, *Theories of Emotion* (pp. 3-33). New York: Academic Press.

Plutchik, R. (2001). The Nature of Emotions. *American Scientist, 89*, 344 - 350.

Rigg, M. G. (1937). Musical expression: An investigation of the theories of Erich Sorantin. *Journal of Experimental Psychology, 21*, 442-55.

Scherer, K. R. (1995). Expression of Emotion in Voice and Music. *Journal of Voice, 9*(3), 235-248.

Shand, A. F. (1914). *The Foundations Of Character – Being a study of the Emotions and Sentiments.* Macmillan.

Thompson, W. F., & Robitaille, B. (1992). Can composers express emotions through music? *Empirical Studies of the Arts, 10*, 79-89.

Trainor, H. (2003). Games localization: production and testing. *Multilingual Computing & Technology, 14*(5), 17-20.

Walker, R. (1996). Open peer commentary: Can we understand the music of another culture? . *Psychology of Music, 24*, 103-130.

*Table 1.*Totals for Tempo Modifications by Emotion and Calculated Chi Square Values

Tempo Measure	Emotion								
	Joy	Sadness	Anger	Fear	Acceptance	Disgust	Surprise	Anticipation	No Emotion (Control)
	Tempo Change								
Accelerating	0	0	4	2	0	1	2	2	0
Decelerating	0	4	0	1	0	1	0	0	0
Constant	9	5	5	6	9	7	7	7	9
X^2	18.00	4.67	4.67	4.67	18.00	8.00	8.67	8.67	18.00
	Tempo								
Faster	8	0	7	4	2	3	8	6	0
Slower	0	9	1	3	6	4	1	2	7
No Change	1	0	1	2	1	2	0	1	2
X^2	12.67	18.00	8.00	0.67	4.67	0.67	12.67	4.67	8.67

Note. $X^2_{.05}$ (2) crit = 5.99

Table 2. Totals for Pitch Modifications by Emotion and Calculated Chi Square Values

Pitch	Emotion								
	Joy	Sadness	Anger	Fear	Acceptance	Disgust	Surprise	Anticipation	No Emotion (Control)
Up	6	0	0	5	3	1	8	6	1
Down	2	9	9	3	3	8	1	1	3
No Change	1	0	0	1	3	0	0	2	5
X^2	4.67	18.00	18.00	2.67	0.00	12.67	12.67	4.67	2.67

Note. $X^2_{.05}$ (2) crit = 5.99

Table 3. Totals for Rhythm Modifications by Emotion and Calculated Chi Square Values

Rhythm	Emotion							
	Joy	Sadness	Anger	Fear	Acceptance	Disgust	Surprise	Anticipation	No Emotion (Control)
Regular	2	5	2	3	4	0	0	5	8
Triplet	3	0	0	0	1	0	0	1	0
Dotted	3	1	0	0	3	1	0	0	0
Double-dotted	1	0	4	3	0	1	4	1	0
Syncopated	0	1	1	2	1	1	3	1	0
Irregular	0	2	2	1	0	6	2	1	1
X^2	6.37	11.73	7.71	6.37	9.05	17.09	10.39	10.39	34.52

Note. $X^2_{.05}$ (5) crit = 11.07

Table 4. Totals for Timbre Modifications by Emotion and Calculated Chi Square Values

Timbre	Emotion								
	Joy	Sadness	Anger	Fear	Acceptance	Disgust	Surprise	Anticipation	No Emotion (Control)
Piano	2	0	3	0	3	2	4	4	8
Celeste	5	0	0	0	2	0	3	2	0
Organ	0	1	0	2	0	1	0	0	0
Guitar	0	0	0	0	1	0	0	0	0
Violin	0	5	0	2	0	0	0	1	0
Choir	0	0	1	5	0	0	1	0	0
Trumpet	0	1	0	0	1	0	0	0	0
Sax	0	0	2	0	0	0	0	0	0
Oboe	0	1	1	0	1	2	0	0	0
Flute	2	1	0	0	1	1	1	2	0
Synth	0	0	2	0	0	3	0	0	1
X^2	31.34	26.45	14.23	31.34	11.78	14.23	24.01	21.56	70.46

Note. $X^2_{.05}$ (10) crit = 19.68

Table 5. Totals for Harmony Modifications by Emotion and Calculated Chi Square Values

Harmony	Emotion								
	Joy	Sadness	Anger	Fear	Acceptance	Disgust	Surprise	Anticipation	No Emotion (Control)
Fifth	3	4	5	5	2	7	3	3	0
None	6	5	4	4	7	2	6	6	9
X^2	1.00	0.11	0.11	0.11	2.78	2.78	1.00	1.00	9.00

Note. $X^2_{.05}$ (1) crit = 3.84

Table 6. Totals for Accompaniment Modifications by Emotion and Calculated Chi Square Values

	Emotion								
Accompani-ment	Joy	Sadness	Anger	Fear	Acceptance	Disgust	Surprise	Anticipation	No Emotion (Control)
Bass									
Triads	0	0	4	2	0	1	2	2	0
Arp. Triads	0	4	0	1	0	1	0	0	0
Tonic Bass	0	4	0	1	0	1	0	0	0
None	9	5	5	6	9	7	7	7	9
X^2	18.00	4.67	4.67	4.67	18.00	8.00	8.67	8.67	18.00

Note. $X^2_{.05}$ (3) crit = 7.81

Table 7. Totals for Dynamics Modifications by Emotion and Calculated Chi Square Values

Dynamics	Emotion								
	Joy	Sadness	Anger	Fear	Acceptance	Disgust	Surprise	Anticipation	No Emotion (Control)
Very Quiet	0	3	0	0	0	0	0	0	1
Quiet	2	4	0	3	0	0	0	0	1
Moderate	2	0	0	1	8	2	3	7	7
Loud	4	0	0	2	1	1	2	0	0
Very Loud	1	0	6	0	0	5	3	0	0
Getting Louder	0	0	3	3	0	1	1	2	0
Getting Quieter	0	2	0	0	0	0	0	0	0
X^2	10.49	13.62	26.12	8.93	41.74	15.18	8.93	32.37	30.80

Note. $X^2_{.05}$ (6) crit = 12.59

Table 8. Totals for Drum Rhythm Modifications by Emotion and Calculated Chi Square Values

Drum Rhythm	Emotion								
	Joy	Sadness	Anger	Fear	Acceptance	Disgust	Surprise	Anticipation	No Emotion (Control)
1	0	5	2	1	1	5	2	2	0
2	1	0	4	6	2	1	2	3	0
3	5	0	1	0	2	2	1	1	0
4	2	0	1	0	2	1	0	1	0
None	1	4	1	2	2	0	4	2	9
X^2	8.22	13.78	3.78	13.78	0.44	8.22	4.89	1.56	36.00

Note. $X^2_{.05}$ (4) crit = 9.49

Table 9. Totals for Attack Length Modifications by Emotion and Calculated Chi Square Values

Attack Length	Emotion								
	Joy	Sadness	Anger	Fear	Acceptance	Disgust	Surprise	Anticipation	No Emotion (Control)
Short	5	0	3	2	2	3	7	2	2
Medium	1	0	4	6	4	4	1	5	5
Long	3	9	2	1	3	2	1	2	2
X^2	2.67	18.00	0.67	4.67	0.67	0.67	8.00	2.00	2.00

Note. $X^2_{.05}$ (2) crit = 5.99

Table 10. Totals for Articulation Modifications by Emotion and Calculated Chi Square Values

Articulation	Emotion								
	Joy	Sadness	Anger	Fear	Acceptance	Disgust	Surprise	Anticipation	No Emotion (Control)
Even	3	7	1	4	2	4	4	7	8
Beat-stressed	5	0	6	2	6	0	1	0	1
Random	1	2	2	3	1	5	4	2	0
X^2	2.67	8.67	4.67	0.67	4.67	4.67	2.00	8.67	12.67

Note. $X^2_{.05}$ (2) crit = 5.99

Table 11. Totals for Scale Modifications by Emotion and Calculated Chi Square Values

Scale	Emotion								
	Joy	Sadness	Anger	Fear	Acceptance	Disgust	Surprise	Anticipation	No Emotion (Control)
Major	8	0	2	1	4	0	6	7	6
Minor	0	9	3	5	0	2	0	1	0
Pentatonic	1	0	1	0	5	2	2	0	1
Blues	0	0	1	1	0	1	0	1	2
Gypsy	0	0	2	2	0	4	1	0	0
X^2	27.11	36.00	1.56	8.22	13.78	4.89	13.78	19.33	13.78

Note. $X^2_{.05}$ (4) crit = 9.49

Table 12. Confusion Matrix of Data for those Familiar with the System

	Predicted									
Emotion	Joy	Sadness	Anger	Fear	Accep-tance	Disgust	Surprise	Antici-pation	Emotion (Control)	% Correct
Actual										
Joy	17	0	0	0	5	0	8	4	0	0.50
Sadness	0	22	2	2	2	2	0	3	0	0.67
Anger	0	0	18	4	0	7	2	1	0	0.56
Fear	0	0	0	16	0	3	2	11	0	0.50
Acceptance	9	0	0	0	7	0	3	6	3	0.25
Disgust	0	0	5	8	0	17	1	0	0	0.55
Surprise	12	0	0	0	3	0	11	2	0	0.39
Anticipation	14	0	0	0	14	0	3	0	0	0.00
No Emotion (Control)	2	0	0	0	3	0	0	3	18	0.69

Table 13. Confusion Matrix of Data for the Independent Test Subjects

Emotion	Predicted									% Correct
	Joy	Sadness	Anger	Fear	Accep-tance	Disgust	Surprise	Antici-pation	Emotion (Control)	
Actual										
Joy	17	0	0	0	3	0	6	0	3	0.59
Sadness	0	27	0	0	0	0	0	0	0	1.00
Anger	0	0	14	8	0	3	0	6	0	0.45
Fear	0	0	6	12	0	3	0	5	3	0.41
Acceptance	3	0	0	0	15	0	3	3	3	0.56
Disgust	0	0	0	9	0	17	0	3	0	0.59
Surprise	3	0	0	0	0	0	17	6	3	0.59
Anticipation	12	0	0	0	6	0	2	3	6	0.10
No Emotion (Control)	0	0	0	0	3	0	0	6	18	0.67