Real-Time Music Generation for a Virtual Environment

Judy Robertson, Andrew de Quincey, Tom Stapleford and Geraint Wiggins
Department of Artificial Intelligence, University of Edinburgh
80 South Bridge, Edinburgh EH1 1HN, Scotland

Abstract

We describe how we are adapting musical techniques used in films to build a computer program to generate atmospheric music suited to an educational virtual environment. The generator produces music to convey fear using suspense and surprise. The paper motivates the search for a mapping between these emotions and musical structure and outlines how the music generator is implemented. It also explains how we intend to evaluate the effects the music has on the users' subjective sense of *presence* or "being there".

1 Introduction

A goal of virtual reality research is to engineer synthetic environments in which users experience a subjective sense of *presence* or "being there". The user's interactions within this virtual world seem to her to be as real as her interactions in the real world. Her behaviour may reflect this through lack of response to real world stimuli and physical response to virtual world events. It is not yet clear exactly which features of a virtual environment might lead users to feel that they are "there inside the environment". Successful virtual environments have achieved this effect through artistry and intuitive design sense rather than through principled physiological and psychological theories. Virtual environment designers can draw on artistic techniques from other media; drawing audiences into a virtual world has long been the goal for novelists, composers, playwrights, and television and cinema directors. A bank of scientific knowledge predicting how visual, auditory and haptic technology will influence users' physical and mental responses to a virtual environment would be complementary to the artistic techniques and might explain why some of the techniques are so effective. Virtual reality researchers working towards a theory of presence are gradually building up this bank of knowledge. Current thinking on presence is explored in section 4.

The aim of our research is to use music to add to the illusion that users are really present within our educational virtual environment, *GhostWriter*. Based on our analysis of the musical techniques used to evoke fear and suspense in films, we have implemented a computer program to generate music dynamically according to user controlled parameters representing increases and decreases in suspense. Our work includes both the artistic and scientific aspects discussed above. We have used artistic knowledge from film and music to include atmospheric music within a virtual environment; the next stage is to analyse the effect this music has on users within the environment and integrate the results of these experiments with existing work on presence.

In this paper, we describe on-going work, full results of which are not yet available – the emphasis is on explaining the motivation for what is a very new area. We focus on this topic in the section 4, and then outline the overall design of the Ghostwriter system in section 3. Section 4 discusses some related work on presence, and section 5 explains how we wish to use music to enhance presence in

GhostWriter. We then outline the design of our system, in section 6, and of our intended evaluation, in section 7. We conclude in section 8.

2 Motivation

Synthetic environments have been used in education for some time. Since Papert's influential work on Logo (1980), there has been interest in educational microworlds for teaching a range of subjects including mathematics, physics and foreign language teaching. There is great potential to improve the educational effectiveness of these applications by taking advantage of advances in computing technology to improve display, input and interactivity. Theories of presence suggest that these environmental features can be manipulated to heighten or lessen the feeling that the learner is really present in the environment. Presence has been linked to motivation – the more a learner is involved in the virtual world, the more she feels as if she is there, and the more motivating the experience will be. In a large scale study of a multimedia synthetic environment for teaching biology, Lester *et al.* (1997) found that the participants remembered more about the lesson when it involved interactions with an engaging animated, talkative synthetic character, compared with a control situation where the material was presented without this character. These results can be explained by reference to theories of perceptual and social presence as discussed in the next section. These early results show that a sense of presence in an educational virtual environment can have important motivational and cognitive effects on the learner.

GhostWriter is an educational virtual environment for supporting the development of children's writing skills. It is a perceptually realistic, highly interactive program in which children interact with each other to improvise a dialogue for a scene in a play. The improvisation task and subsequent writing tasks based on it are intended to: develop expressive and poetic writing skills; encourage children to write for a wide range of audiences; and to develop affective maturity through writing. The latter educational aim of GhostWriter is particularly dependent on evoking a sense of presence in learners.

Affective maturity is both displayed in and developed by writing tasks (Wilkinson, 1983). Writing can be categorised by several dimensions of affective maturity, including how far the writer portrays and analyses their own and other characters' emotions and the extent to which they consider the environment for the story and how it influences the characters. The hypothesis is that these aspects of affective maturity can be developed through experiences in a virtual environment. More specifically, the former may be developed through interactions with the other users – social presence; the latter may be developed through interactions with the virtual environment itself – perceptual presence.

We expect that atmospheric dynamically generated music will alter the users' emotional states during the improvisation in the virtual environment (an aspect of presence of self). Our hypothesis is that this will develop the writers' skills at portraying believable emotional reactions for characters and therefore develop affective maturity.

3 Ghostwriter

Although the approach used in GhostWriter could be used to develop writing skills in any genre, the current implementation is tailored towards horror. It is expected that the writing skills developed through use with the system will be generally beneficial but that concentrating on one genre would render the development task more tractable. Horror seemed appropriate because librarians and teachers report that ghost and horror stories are extremely popular with children between the ages of eight

and sixteen, and both the familiarity and popularity of the genre would be motivating. For these reasons, GhostWriter is a computer based haunted house. It is a 3D computer graphics version of an old manor house, containing all the usual sorts of things to be found in haunted houses – cellars, attics, haunted baths, secret passageways and dog eared family bibles. Users can interact within this world, by walking from room to room, and picking up ornaments and reading books. As they wander about in the haunted house they can hear strange creakings, clocks ticking and possibly eerie howls. Atmospheric music plays in the background. They can also interact with other users by typing messages to them. The users in the haunted house are represented by animated characters (avatars). Each avatar is controlled by a user, who can choose a facial expression for the avatar to depict their current mood. Like current 3D computer games, this environment will run on home PCs.

What is described so far would make a mildly entertaining computer game; the following details show why GhostWriter is an educational environment. A typical session within the environment would proceed as follows. A group of children and their teacher would decide to act out a scene of a play set in a haunted house. After talking about it for a while, and deciding who should play what character, each group member could move to a computer and use the friendly interface character in GhostWriter to help them plan out their character in more detail. Part of planning the characters involves planning what the character's goal should be for that scene. The ghost character might plan to take terrible revenge on the others, while the scientist might set out to prove that ghosts do not exist.

Once each character was thought out, the teacher would lead a role playing session within the virtual haunted house. This would be a dialogue between all the group members as they interacted in the scene. The scene could take place in any of the rooms of the haunted house, and could move from room to room. As the group members moved their avatars around the scene they would type messages to the other characters to add to the dialogue. They could also choose facial expressions for the avatar to match what the character would be feeling. The teacher could take on the role of director by cueing the characters to speak, or perhaps play a character in the scene. It is at this stage that dynamically generated music is important. The director can use music to influence the atmosphere of the scene, and convey appropriate emotions to the role players at certain parts of the scene. For example, when the ghost first appears, the director might want to prompt the role players to act scared and would adjust the parameters to the music generator so that it would output scarier music.

As the group members improvise via typing their lines, GhostWriter stores the dialogue so far. At the end of the improvisation session, each group member can have a copy of what was said in the scene. This script can be edited in a word processor (by each member of the group separately) to improve and extend it.

At the end of the scene, the teacher would lead a debriefing discussion in which the group members exchange their experiences of "being" their character, how it felt, and how it could be used in a story. GhostWriter could be used to create a whole play with several scenes in successive sessions with the same group, or it could be used on one off occasions to create just one scene. The same scene could be improvised again, until the group members are satisfied with it.

4 Background

This section discusses research into presence and the aspects of presence which are important in GhostWriter and the atmospheric music generator. Intuitively, presence refers to a sense of being there – being *part of* a virtual environment, while using it. However, this is too loose a definition to be useful as a basis for scientific investigation. A stricter definition comes from Lombard and Ditton (1997): they define presence as "the perceptual illusion of non-mediation". When a person using a

communication environment uses her normal perceptual, cognitive and affective systems to respond to the environment as if the experience is not transmitted through a medium, she is said to experience presence. Presence can occur during the use of any of the human made media – television, telephone, cinema or virtual environments. It cannot be felt in degrees; at any instant a media user will either experience presence or not experience it. The sense that some media heighten or lessen presence stems from the ratio of instants with presence to instants without presence during media use.

A number of factors contribute to an overall sense of presence – physical presence, social presence and the presence of self.

Presence is affected by the perceptual realism of the virtual world portrayed using the medium – how much does a mediated experience look, sound, feel, taste and smell like an unmediated experience? When the medium stimulates more of the user's senses, the more the user is immersed in the system, and so presence is more likely during the experience. Visual and auditory stimuli are skillfully manipulated in film, television, and to a lesser extent theatre, to increase the realism of the production. There is a large body of knowledge about the technical and artistic techniques for creating presence in these media, and some research has been conducted into how varying aspects of a visual display medium can alter the sense of presence (reported by Lombard and Ditton, 1997). There has also been some research into how auditory display affects presence. Anderson and Casey (1997) distinguish between three types of audio display – live speech, Foley sounds and ambient sounds. Live speech is used in real time communication and is most likely to affect social presence. Ambient sounds, including music, are most likely to alter the users' emotional states and therefore alter the perception of presence of self. Foley sounds are sound effects triggered by objects in the environment and contribute to perceptual presence. For example, the sound of water lapping audible beside a lake in a virtual environment is a Foley sound. Considerable research effort has been concentrated on including realistic Foley sounds in an environment, and on giving the sound the appropriate acoustics and directionality.

Perceptual presence is required in GhostWriter to give users a sense of being physically present in an environment which they have never visited, and in which ghostly events can happen. This experience is intended to develop affective maturity in the users' writing because it will highlight the relationship between an environment and the characters acting in it. Perceptual presence will be heightened by highly interactive environment, with almost photo realistic 3D graphics and a large selection of Foley sounds.

A definition of *social presence* is the "extent to which a medium is perceived as sociable, warm, sensitive, personal or intimate when it is used to interact with other people." (Lombard and Ditton, 1997). Social interactivity within a virtual environment is likely to lessen the feeling that the environment is mediated. Heeter (1992) says "placing more than one person in a virtual world may be an easy way to induce a sense of presence regardless of the other perceptual features of the world". From an audio perspective, social presence could be improved by the use of live speech transmitted between users. However, the technology to support transcription of live speech is not mature enough to make this feasible, and in addition spoken improvisation may not achieve the educational goals as well as written improvisation.

Social presence in GhostWriter is required to promote the development of affective maturity in writing. Interactions with other people will help learners to see other people's view points, and to consider their own more carefully – aspects of maturity towards self and others. Social presence will be heightened in GhostWriter through interactions with other users and with synthetic characters; Biocca and Levy (1995, pages 259-276) suggest that the ability to interact with large numbers of people, for example via multiplayer virtual reality games may increase social presence.

Clearly individual differences alter presence and also the mood and state of the individual before after and during the experience. *Self-presence* is defined as "the users' mental models of themselves

inside the virtual world, but especially differences in self-presence due to the short term or long term effect of virtual environments on the perceptions of one's body (*i.e.*, schema or body image), physiological states, emotional states, perceived traits, and identity" (Biocca, 1997).

Biocca (1997) states that "designers share the assumption that increases in self-presence are correlated with higher levels of cognitive performance, and possibly emotional development."

A medium can evoke emotional responses from the user during the mediated experience. The more emotionally involving the experience is, the more it is likely to change the emotional state of the user and so increase self presence. The mediated experience of film watching can create the illusion of non-mediation for long periods of time by playing on the emotions of the viewer. Some genres of film rely on this – "weepies", comedies and horror films. Music has a particularly important role in evoking emotion in film viewers. Aaron Copland, as reported by Prendergast (1990, pages 215-216), considered that "music can create a convincing atmosphere of time and place" and that "music can be used to underline or create psychological refinements – the unspoken thoughts of a character or the unseen implications of a situation."

The famous film theoretician, Hanns Eisler, and his co-author Adorno, wrote that standard film techniques blend many cinematic arts, including music, to hide the mediated, remote nature of the medium (reported by Brown, 1994). This suggests that music also has the potential to heighten presence (the illusion of non-mediation) in virtual environments.

This is an intuitive viewpoint from an accomplished artist; there have also been scientific studies into the relationship between music and emotion. Krumhansl (1997) discusses the distinction between music evoking and expressing emotion. If music can actually evoke emotion, then listeners would both be able to identify an emotion associated with the music, and would undergo physical changes associated with that emotional state. If music merely expresses emotion, then listeners find that music suggests an emotion to them but they don't experience the emotion through listening. Her results suggest that, since music can cause physiological changes in listeners which are consistent with the emotions they report, music can evoke emotional states. However, for our purposes the fact that listeners identify bits of music as "scary", "happy" or "sad" and can do so consistently, is sufficient – we need only the *connotation* of emotion, and not necessarily the emotion itself.

5 Music in ghostwriter

Music can be used to convey emotion, but at the current time, the relationship between emotion and musical structure is not clear. Much research would be required to design a music generator covering the full spectrum of human emotion; thus we have decided to focus on fear in the first instance. Fear is an appropriate emotion for users to encounter in a virtual haunted house, and writers in the horror genre should become familiar with manipulating the level of fear their readers experience. The music generator will convey fear through manipulating dramatic tension using both suspense and surprise. It is anticipated that the music generator could be adapted in the future to convey other emotions, just as the GhostWriter itself could be adapted to support writing development specific to other fictional genres.

Fear is related to, but is not synonymous with, tension. Griffiths (1982) writes that effective plays manipulate dramatic tension to keep the audience's interest. The playwright should engender, maintain, suspend, heighten and resolve a state of tension in order to make the audiences ask questions like "what is happening?" or "what will happen next?". Naturally, novelists also manipulate dramatic tension – in fact, crime fiction and authors of detective stories often do so to great effect (*e.g.*, "The Spy Who Came in from the Cold" by John Le Carré). According to Griffiths (1982), dramatic tension

can take many forms including: emotional tension between characters as exhibited in a dialogue, erotic tension, hostility or a sense of time running out. Two forms of tension typically found in horror and thriller genres are surprise and suspense. In these forms, dramatic tension is manipulated to make the audience more or less scared; it is used to evoke fear. We have decided, therefore, to concentrate on using surprise and suspense to convey fear through music, by drawing on some of these established techniques in horror and thriller films,

The distinction between suspense and surprise lies in either feeding or confounding the audience's expectations. Suspense builds up tension gradually. The author leads the audience to expect a terrible happening and their anticipation of that event engenders fear and unease. This fear can be heightened by delaying the event for just a little bit longer. It can also be effective to suspend the tension for a bit to lull the audience into a false sense of security by seemingly preventing the terrible event. The event, when it does finally happen, may resolve the tension, and change the audience's fear into another emotion such as sadness. It could also heighten the suspense by predicting another chain of terrible events.

Surprise introduces an unforeseen sudden shocking event into a scene, contrary to the expectations of the audience. This event will induce fear in the audience through a shock, and create tension by causing them to wonder why it happened and what the consequences will be. Surprise may lead on to suspense: *e.g.*, a building is unexpectedly bombed leaving the hero trapped under a fallen piece of masonry. The audience wonder whether he will be rescued before he dies of smoke asphyxiation. There is a difference between surprise and the climax of suspense, although both may be shocking.

Surprise has the shock of the unexpected, while the building of suspense finally to peak at the terrible event shocks the already strained nerves of the audience. Surprise can also be used within suspenseful scenes, perhaps to temporarily reduce tension: *e.g.*, as the hero opens the closet in which the audience firmly believes that a murderer lurks, a trapped cat jumps out.

The above discussion of surprise and suspense highlights some requirements of the music generator. It should be capable of dynamically conveying increases and decreases in suspense, temporary relaxation of suspense, the climax of a suspenseful build up and resolution of suspense. It should also be able to convey surprise during a suspenseful build up and at other points in the scene.

The suspense and surprise parameters to the generator will be manipulated in real time by the director of the improvisation within GhostWriter.

6 Dynamic Generation of Music to Convey Feelings of Fear

6.1 Techniques used in Film Music

Music is often used in films to involve an audience emotionally. Members of the audience can follow the story line as the visuals unfold, and can *understand* the emotions of the characters. However, as Brown (1994) says, it is the combination of the visuals with music that makes the viewers *feel* those emotions. For example, when watching *Psycho* we not only attribute fear to the unfortunate heroine when we watch her murder; we also feel fear, a sensation heightened by the visual impact of the camera work and the shocking musical accompaniment. Indeed, Brown (1994) writes that a critic for Time magazine objected violently to this scene because he considered it gruesome and far too explicit. On closer examination of the film, though, the portrayal of the murder does not actually show the knife entering the victim's body. What causes the horror is the combination of Herrman's score and Hitchcock's macabre genius in editing the camera shots.

Music is used in two modes within films: diagetic and non-diagetic. Diagetic music is music which occurs within the narrative of the film. It is attributable to some source seen in the film, such

as a radio, a record player, a musical instrument, a juke box or an orchestra in a concert hall. The characters of the film can hear this music. Non-diagetic music is music which is not part of the narrative, not attributable to a source in the film, and is unheard by characters in the film. This type of music is usually used to add affective colour to the film, and play on the emotions of the audience. It can be used to cue the audiences to feel uneasy; to build up tension; to flag an approaching disaster or to signal a love affair. Some composers, such as Korngold (composer of the music for the 1940 film "Sea Hawk") associated particular characters with a non-diagetic musical theme, in the leitmotivic style (see below) most commonly attributed to Wagnerian Opera. This type of music is, for the most part, used to cue events in the narrative, although non-diagetic music playing at a counterpoint to the narrative can achieve interesting results.

Our music generator will produce music which lies somewhere between these categories. Although the music it produces will not have a source in the virtual environment, and will be used principally to evoke emotion, it will be nonetheless heard by the characters in the virtual environment. This twist results from the unusual situation in which the audience are also participants in the narrative. GhostWriter will also use Foley sounds for diagetic cues in the environment, such as creaky doors or crackling fires.

The task of composing music suited to film is not analogous to composing concert hall music. Film composers have the disadvantage that the audience are not giving their full attention to the music because they are also watching the visuals; also, the film composer's musical narrative is usually constrained in some ways by the narrative of the film. A composer can be further frustrated by the director's haphazard editing; music can be cut and pasted back together in a way which achieves the director's aims but makes no large-scale musical sense. In addition, music cues are often very short, thus restricting any serious development of coherent musical form. For this reason, film music often consists of elaborations and extensions of a particular short motif, or a theme associated with a character (a *leitmotif*). For example, in his work with Hitchcock, Herrmann solved these problems by reducing the role of melody and relying more on short-term harmonic and rhythmic constructions. Because of these artificial restrictions, film-style music is a genre more readily accessible to computer composition than concert-hall music: the scope of compositional vision is smaller, and so the complexity of the compositional task is reduced. The practice of using leitmotifs to herald characters may also be fruitful for the role playing aspect in Ghostwriter.

The music for ghostwriter will be generated live, during the improvisation of the narrative. In contrast to the usual method of composition for films, where the composer can watch the film all the way through to fit the music to the narrative, our music generator will have no access to information about what will happen in the narrative. The level of dramatic tension, as indicated by the director has to change dynamically with no indication of future events. This method of improvising is more related to the method used by the director Louis Malle for his 1957 film, "Ascenseur pour l'echfaud" when he recorded a jazz quintet to improvising music as they watched the film in the studio.

There are many techniques for creating dramatic tension in music. (Note that the concept of *musical tension* is distinct from that of dramatic tension, though in practice the two often coincide.) For example, increasing degrees of *dissonance* in music, broadly characterised as the number of small intervals occurring in chords, or the number of notes not in the current key used in a given passage, often create increasing dramatic tension. Colloquially, dissonant music is often described as "more jarring" than consonant music.

Different modes (that is, different kinds of musical scale), notably the common major and minor modes, have very different dramatic implications. Well-known examples of these, respectively, are Vivaldi's "Four Seasons: Spring", with its optimistic brightness, and the third movement of Chopin's Piano Sonata number 3, in B♭ minor, which has come to be known as "The Death March", with its

beautiful but sombre harmony. These characteristics give general moods to the music; a more specific kind of musical manipulation is the generation of active expectation, where musical structures are allowed to develop (or perhaps simply repeat) in some predictable way. Dramatic (and musical) tension can then mount in a gradual build-up, or, more shockingly, in sudden denial of a very strong expectation.

A particularly rich source of musical techniques used to create atmospheres of fear can be found in the scores for the films "Vertigo", "North by NorthWest" and "Psycho". These three films were the most effective results of the Hitchcock and Herrmann collaboration, which is reported in detail by Brown (1994). In order to overcome the standard obstacles to the film-music composer, described above, Herrman took the course of using minimal *melodic* material, and basing his music instead on harmonic and rhythmic textures and motifs. In particular, a chord, which has become known as Herrman's "Hitchcock Chord" was used, which had particular properties. The full musical detail of how the Hitchcock Chord works is not relevant to the present paper; suffice it to say that it is a musical construction which is strongly ambiguous between the major and minor modes, and can be interpreted harmonically in different ways according to context. Thus, Herrman was able to produce a consistent harmonic flow, with the tone, or *timbre* of his chordal work, but he was able to generate different feelings by the context in which he placed it.

Similarly, in "The Trouble with Harry", Herrman's first collaboration with Hitchcock, the Hitchcock Chord is used, but in this much lighter-hearted film, it is understated, and does not appear complete. The partial appearances, however, are linked with an obsessive repetition of a two-bar rhythm, which always appears in groups of five. Again, this is explained in detail by Brown (1994).

Studying Herrman's and others' work in detail gives us a rich source of ideas for how to develop the musical rules which we will use in our music generator.

6.2 The Music Generator

6.2.1 Overview

Our system consists of three main subsystems: the music generator (responsible for actually creating the music), the mapping system (responsible for accepting the user's input of "tension level" and using that to control the generator), and a set of utilities (for loading and configuring the system, etc.)

The generator contains four main modules: the *High-Level Form* generator, the *Rhythm and Volume* generator, the *Melody and Harmony* generator, and the *Player*. The tension level (set by the user) is mapped onto a set of parameters that influence the music generator modules, thereby controlling the tension level of the resulting music.

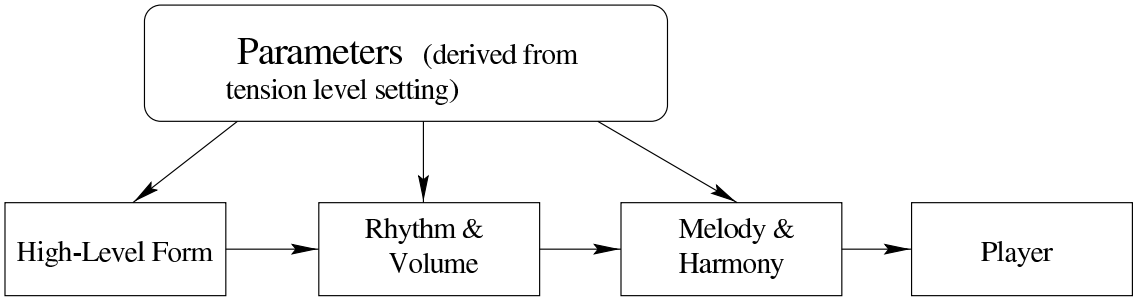

Figure 1: Diagram of Generator Modules

The modules are connected to one another in a linear fashion (see Figure 1) and pass pointers to musical data structures from one module to the next, much like a conveyor belt assembly system. Newly created musical data structures contain uninstantiated attributes (pitch, volume, etc.). Each module is responsible for instantiating certain attributes; as the data structure is passed along, it gradually becomes more and more specific, becoming fully instantiated by the time it reaches the *Player* and is translated into MIDI commands (Rothstein, 1992). Old, previously instantiated data structures may also be drawn from a *Phrase Library* (by the *High-Level Form*), possibly partially uninstantiated, and then passed through the system to receive new attribute instantiations as necessary.

6.2.2 Musical Data Representation

The musical data is represented using CHARM, a hierarchical, abstract musical representation presented by Wiggins et al. (1989) and Smaill et al. (1993). The basic element of CHARM is the *event*, which represents a note of fixed pitch. Events can be grouped together under higher level structures (a chord, for example) called *constituents*. In turn, lower-level of constituents can be further grouped under a higher level constituent (a phrase, for example). The key feature of CHARM needed here is its ability to represent multiple hierarchies, so that orthogonal properties and groupings of a set of musical events can be represented.

6.2.3 Harmony Generation

In order to realise Herrman's techniques (outlined above), it was necessary to develop a fairly structured knowledge base of the rules of tonal Western tonal harmony. Arnold Schoenberg's Theory of Harmony (1984) was an ideal source for these rules, as Schoenberg provides not only an explicit set of guidelines for the student composer that are developed over the course of the text (pages 123-125, for example), but also a theoretical basis that emphasises the manner in which the the rules tend to stabilise or destabilize the tonal focus that forms the heart of tonal music. Thus, Schoenberg provides the methods for implementing Herrman's technique of creating suspenseful tension by breaking down the rules of traditional Western tonal music.

Although Herrman provides the means by which music can be manipulated to create suspenseful tension, and Schoenberg provides the methods to perform those manipulations, there is still the problem of linking the musical developments to the visual/narrative developments. Part of this difficulty is solved in our system by having the teacher/facilitator move a graphical slider up and down to reflect the level of tension that should be present in the music. Still, the tension level must be quantified and related to the parameters of the music generation system. For this first version of the music system, this is accomplished by a hand-crafted set of rules and parameter values that effectively form a mapping between the level of the graphical slider and the level of tension in the music; other methods, such as machine learning-based approaches, are clearly of interest here.

The *Melody and Harmony Module* is responsible for instantiating the *events* and *chords* output by the *Rhythm and Volume* module with pitch (or chord type) information according to this mapping. It works in three steps:

1. Generate a chord type for the available chord structure
2. Instantiate each available melodic event with a pitch
3. Instantiate each available accompaniment event with a pitch (based on the chord type of its associated chord)
4. Assign instrumentation to each available event (both melody and accompaniment)

The Harmony sub-module generates chord progressions containing the four basic triad types (major, minor, augmented, and diminished) and the seventh and ninth chords formed from these triads. The Harmony sub-module selects chords based on a wide variety of parameters including: current musical key (major and minor modes supported on all twelve equal-tempered octave tones), desired level of dissonance, resolution/non-resolution of dissonances, and strong/weak chord progression intervals. The Harmony sub-module is also capable of initiating several types of cadences (perfect, plagal, deceptive, *etc.*) or making smooth modulations between major and minor modes when desired. Final selection of a chord is performed probabilistically, after all of the parameters have influenced the probability function. At the present time, the Harmony sub-module does not consider voice-leading and therefore does not deal with chord inversions.

The Melody sub-module instantiates the *events* that comprise the melody with pitches. Pitch selection is probabilistic, and influenced by a number of factors, including: probability of pitch being a chord tone, current musical key, preferred interval sizes, desired interval harshness, and melodic balance.

6.2.4 Rhythm Generation

Harmony and melody are sometimes thought of, particularly by contemporary composers, as the main aspects of music. In film music, this is clearly not the case, as can be heard, for example, the distinctive menacing rhythms of the "Jaws" soundtrack.

In our system, we have separated out the construction of rhythms from the construction of harmony and melody, for purposes of simplicity. This is clearly an over-simplification in terms of human musical composition, but it is a workable compromise in this context.

Rhythms can be thought of as composed from a set of five rhythmic feet: Iamb, Anapest, Trochee, Dactyl, and Ambibrach, which are more commonly used to describe poetic rhythms (Cooper and Meyer, 1960). These feet are composed of two or three sets of beats, one of which is *accented, i.e.*, distinguished from the others in some way. Different combinations of rhythmic feet can be composed to generate a wide variety of rhythms, and convey different emotional levels.

While this basic theory is well understood, and has been used successfully in AI applications (Steedman, 1973, for example), there seems very little literature on our specific interest. Cooper and Meyer, as mentioned above, cover rhythm in general, and Edmonds (1982) has a chapter discussing rhythmic feet in film music, but there is little else.

To implement a system based on the idea of rhythm feet, it was necessary to find a set of rules describing how to combine rhythmic feet in response to a desired emotional level. These rules were extracted by viewing many films, and taking notes on the rhythmic structure and the corresponding emotional level, on a purely empirical basis. The effects of the rhythmic rules will be tested, along with the harmonic and melodic ones, as outlined in section 7.

7 Evaluation Intentions

Although the generator will be tested with adults during the development phase, we feel that it is most important to test it with the target user group of GhostWriter. The intended user group are children, and it is likely that there are affective differences in the way music influences listeners of different ages. It is also possible that the interaction between music and other features of a virtual environment has different affects on the subjective feeling of presence, depending on the age of the user.

We will use a group of eleven year old members of a local community story telling club for the

evaluation. This is a particularly interesting group from the point of view of evaluating GhostWriter because they are so familiar with oral narrative and have participated in some drama work. They have also recently worked with the Scottish National Orchestra to produce a set of songs and poems for a Burns Night Celebration.

The first stage of the evaluation will be to assess whether the music generator can produce music which test participants describe as "scary". In addition, the participants will be asked to respond if they notice that the music gets gradually scarier, gradually less scary or suddenly very scary. These indicators correspond to suspense and surprise, respectively. The methodology will be to create graphs of dramatic tension, as indicated by turning a knob while listening to the music, by means of a MIDI-equipped computer.

The next stage of the evaluation will be carried out within the GhostWriter environment. Firstly, we are interested in the interplay between music, Foley sounds and visual cues in the environment, and how this alters the users' subjective feelings of presence; this will contribute to knowledge about perceptual presence. Secondly, in an investigation of the interaction between music and social presence, we will examine the affect the music has on the improvised dialogue. Lastly, we will study the relationship between music and presence of self – did the music make the experience more emotionally involving for the users? These studies will take place after the completion of the ghostwriter environment.

8 Conclusion

In this paper, we have presented some on-going work on the automated, real-time production of live music for an educational virtual environment. We have based our musical ideas on those of a master film composer, Bernard Herrman, and a master theoretician, Arnold Schoenberg, and focussed on the application of known musical principles in an AI environment. We have outlined the design of such an environment, which we are, at the time of writing, engaged in implementing.

In future, we expect to evaluate our system, as described above, and to feed back the results into further developments.

Acknowledgements

This research is supported by a US National Science Foundation postgraduate study grant to Tom Stapleford, EPSRC PhD Studentship 97305815 to Judy Robertson and EPRSC Advanced Course Studentship 97401162 to Andrew de Quincey.

References

Anderson, D. and Casey, M. (1997). The sound dimension. *IEEE Spectrum*, 34(4):46–51.

Biocca, F. (1997). The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of Computer Mediated Communication*, 3(2). Available online at `http://www.ascusc.org/jcmc/vol3/issue2/lombard.html`.

Biocca, F. and Levy, M. (1995). *Communication in the age of virtual reality*. Lawrence Erlbaum, Hillsdale, NJ.

Brown, R. S. (1994). *Overtones and Undertones.* University of California Press, Berkeley and Los Angeles, CA.

Cooper, G. and Meyer, L. (1960). *The Rhythmic Structure of Music.* University of Chicago Press, Chicago.

Edmonds, R. (1982). *The sights and sounds of cinema and television : how the aesthetic experience influences our feelings.* Teachers College, Columbia University, New York.

Griffiths, S. (1982). *How Plays Are Made.* Heinemann Educational Books, Oxford.

Heeter, C. (1992). Being there: The subjective experience of Presence. *Presence*, 1(2):262–271.

Krumhansl, C. (1997). Musical tension: Cognitive, motional and emotional aspects. In *Proceedings of the 3rd Triennial ESCOM Conference.* European Society for the Cognitive Sciences of Music.

Lester, J., S., C., Stone, B., Kahler, S., and Barlow, S. (1997). Animated pedagogical agents and problem solving effectiveness: A large-scale empirical evaluation. In Du Boulay, B. and Mizoguchi, R., editors, *Artificial Intelligence in Education: Proceedings of AI-ED 97*, pages 615–617, Amsterdam. IOS Press.

Lombard, M. and Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer Mediated Communication*, 3(2). Available from http://www.ascusc.org/jcmc/vol3/-issue2/lombard.html.

Papert, S. (1980). *Mindstorms: Children, computers and powerful ideas.* Harvester, Brighton.

Prendergast, R. (1990). *Film Music: A Neglected Art.* Norton and Company, New York.

Rothstein, J. (1992). *MIDI : a comprehensive introduction.* Oxford University Press, Oxford.

Schoenberg, A. (1978). *Theory of Harmony.* Belmont Music Publishers, 3rd edition.

Smaill, A., Wiggins, G., and Harris, M. (1993). Hierarchical music representation for analysis and composition. *Computers and the Humanities*, 27:7–17. Also from Edinburgh as DAI Research Paper No. 511.

Steedman, M. (1973). *The Formal Description of Musical Perception.* PhD thesis, University of Edinburgh.

Wiggins, G., Harris, M., and Smaill, A. (1989). Representing music for analysis and composition. In Balaban, M., Ebcioglu, K., Laske, O., Lischka, C., and Sorisio, L., editors, *Proceedings of the 2nd IJCAI AI/Music Workshop*, pages 63–71, Detroit, Michigan. Also from Edinburgh as DAI Research Paper No. 504.

Wilkinson, A. (1983). Assessing language development: the crediton project. In Freedman, A., Pringle, I., and Yalden, J., editors, *Learning to Write: First Language/Second Language*. Longman, New York.