1-1-2010

Emotion, Content & Context in Sound and Music

Stuart Cunningham
Glyndwr University, s.cunningham@glyndwr.ac.uk

Vic Grout
Glyndwr University, v.grout@glyndwr.ac.uk

Rich Picking
Glyndwr University, Wrexham, r.picking@glyndwr.ac.uk

Recommended Citation
Cunningham, S., Grout, V. & Picking, R.(2010). 'Emotion, Content & Context in Sound and Music', In Grimshaw, M. (Ed), Game Sound Technology and Player Interaction: Concepts and Developments. IGI Global Publishing. pp. 235 - 263.

Emotion, Content & Context in Sound and Music

Abstract
Computer game sound is particularly dependent upon the use of both sound artefacts and music. Sound and music are media rich in information. Audio and music processing can be approached from a range of perspectives which may or may not consider the meaning and purpose of this information. Computer music and digital audio are being advanced through investigations into emotion, content analysis, and context, and this chapter attempts to highlight the value of considering the information content present in sound, the context of the user being exposed to the sound, and the emotional reactions and interactions that are possible between the user and game sound. We demonstrate that by analysing the information present within media and considering the applications and purpose of a particular type of information, developers can improve user experiences and reduce overheads while creating more suitable, efficient applications. Some illustrated examples of our research projects that employ these theories are provided. Although the examples of research and development applications are not always examples from computer game sound, they can be related back to computer games. We aim to stimulate the reader's imagination and thought in these areas, rather than attempt to drive the reader down one particular path.

Keywords
content, context, emotional interaction, emotional reaction, emotional state, playlist generation

Disciplines
Computer and Systems Architecture | Digital Communications and Networking | Hardware Systems | Music | Other Computer Engineering | Systems and Communications

Comments
This chapter appears in Game Sound Technology and Player Interaction: Concepts and Developments edited by Grimshaw, M. Copyright 2010, IGI Global, www.igi-global.com. Posted by permission of the publisher. Chapter 12 on Emotion, Content & Context in Sound and Music by Cunningham, S., Grout, V. & Picking, R. Details of the chapter and link to the book are located on the publishers website at http://www.igi-global.com or via the DOI at http://dx.doi.org/10.4018/978-1-61692-828-5.ch012

Game Sound Technology and Player Interaction:
Concepts and Developments

Mark Grimshaw
University of Bolton, UK

INFORMATION SCIENCE REFERENCE
Hershey · New York

Director of Editorial Content:	Kristin Klinger
Director of Book Publications:	Julia Mosemann
Acquisitions Editor:	Lindsay Johnston
Development Editor:	Joel Gamon
Publishing Assistant:	Milan Vracarich Jr.
Typesetter:	Natalie Pronio
Production Editor:	Jamie Snavely
Cover Design:	Lisa Tosheff

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Game sound technology and player interaction : concepts and development / Mark Grimshaw, editor. p. cm.
Summary: "This book researches both how game sound affects a player psychologically, emotionally, and physiologically, and how this relationship itself impacts the design of computer game sound and the development of technology"-- Provided by publisher. Includes bibliographical references and index. ISBN 978-1-61692-828-5 (hardcover) -- ISBN 978-1-61692-830-8 (ebook) 1. Computer games--Design. 2. Sound--Psychological aspects. 3. Sound--Physiological effect. 4. Human-computer interaction. I. Grimshaw, Mark, 1963-
QA76.76.C672G366 2011
794.8'1536--dc22

2010035721

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Chapter 12

Emotion, Content, and Context in Sound and Music

Stuart Cunningham
Glyndŵr University, UK

Vic Grout
Glyndŵr University, UK

Richard Picking
Glyndŵr University, UK

ABSTRACT

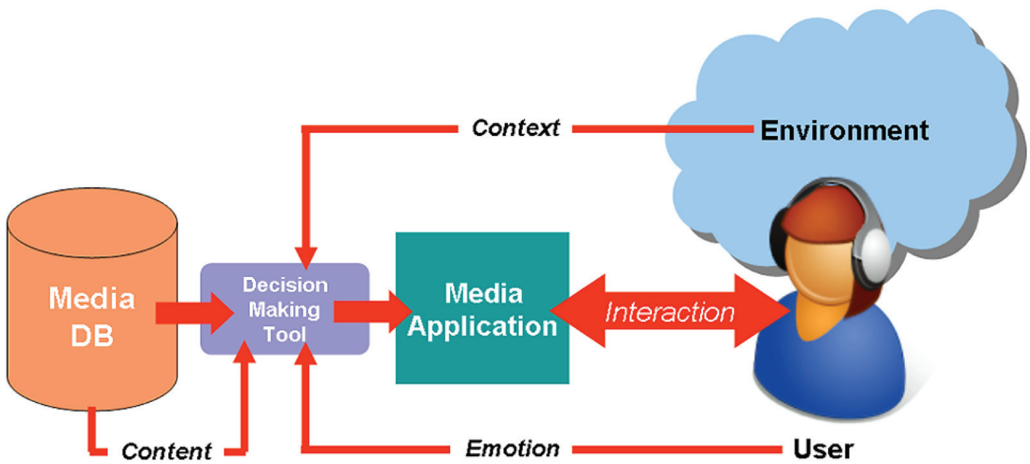
Computer game sound is particularly dependent upon the use of both sound artefacts and music. Sound and music are media rich in information. Audio and music processing can be approached from a range of perspectives which may or may not consider the meaning and purpose of this information.Computer music and digital audio are being advanced through investigations into emotion, content analysis, and context, and this chapter attempts to highlight the value of considering the information content present in sound, the context of the user being exposed to the sound, and the emotional reactions and interactions that are possible between the user and game sound. We demonstrate that by analysing the information present within media and considering the applications and purpose of a particular type of information, developers can improve user experiences and reduce overheads while creating more suitable, efficient applications. Some illustrated examples of our research projects that employ these theories are provided. Although the examples of research and development applications are not always examples from computer game sound, they can be related back to computer games. We aim to stimulate the reader's imagination and thought in these areas, rather than attempt to drive the reader down one particular path.

INTRODUCTION

Music and sound stimulate one of the five human senses: hearing. Any form of stimulation is subject to psychological interpretation by the individual and a cause-and-effect relationship occurs. Whilst this relationship is unique to each individual up to a point, it is safe to assume that broad, often shared, experiences occur across multiple listeners. It can be argued that the emotional reaction and response of a listener to a sound or piece of

DOI: 10.4018/978-1-61692-828-5.ch012

Figure 1. Idealised Role of Emotion, Content and Context in a Computer Application



music is the single most important event resulting from that experience.

The goal of this chapter is to explore the relationship between sound stimuli and human emotion. In particular, this chapter examines the role sound plays in *conveying* emotional information, even from sources that may be visual in origin. Equally, the chapter seeks to demonstrate how human emotion is able to flip this paradigm and influence music and sound selection, based on *emotional state* and consideration of the *context* of the user.

The *content* being represented digitally provides the opportunity to gain a greater understanding of the information present in a data set. Information being stored often has a number of characteristic features and structural elements that can be identified automatically. For example, music generally contains an identifiable structure, which might consist of several movements, parts, or, more commonly, verses and choruses. However, such structure can almost be considered *fractal*, in that there are microscopic and macroscopic levels of organisation and also repetition, ranging from musical beats, bars, verses and choruses to the level of the song itself.

Contextual data provides additional information about factors that contribute to making the user interaction experience much more relevant and effective by acquiring knowledge of the external factors that influence decision making and the emotion of the user.

The conceptual diagram of Figure 1 shows an idealised situation in which a large database of audio media is presented to the user through a suitable application (such as a computer game). In this scenario, the user's *emotion* and *context* are analysed and compared against analysis of appropriate media *content*. This provides selection of the 'best fit' media that will further stimulate and engage the user in the most effective way.

The chapter explains the fundamentals of emotional stimulation using sounds and music, whilst retaining relevance to the audiologist. We demonstrate that by analysing the information present within media and considering its applications, significant advantages can be gained which improve user experiences, reduce overheads, and aid in the development of more suitable, efficient applications: whether they be computer games or other audio tools.

EMOTION

Emotion is a key factor to consider in computer applications given that almost all applications will have some form of Human Computer Interface (HCI). Humans are emotional beings and the interaction with the machine will have some emotional effect on them to a greater or lesser extent. The computer, therefore, has an ability to invoke an emotional response in the user. The user may bring their own emotions with them to an interactive experience which has been affected by external factors in the environment around them (Dix, Finlay, Abowd, & Beale, 2003). The quality and resultant experience that a user has with a machine is important and this is also true when we consider the frequent interaction that we have with entertainment media and computer games.

Emotion in Multimedia

The use of sound in multimedia, and especially in computer games, is commonplace. This is unsurprising if one considers that, in order to successfully engage a human user in an immersive experience, the interaction must be achieved through one of the primary human senses. Speech and hearing are hugely important in our daily lives and allow us as humans to send and receive large amounts of information on an ad-hoc basis.

Naturally, it is hearing and the use of sound that we are interested in examining in this chapter. Sound is used in complementing and augmenting other stimuli, especially visual. Consider, for example, the last time you watched a horror movie and were embarrassed by the unintended jump or flinch you experienced at a big bang or crescendo that accompanied the appearance of the bad guy in the movie! Proof, if it were needed, that the constructive use of music and sound can provoke one of the most primal of human emotional instincts; fear. Sound in multimedia environments is classified into two distinct categories (see Jørgensen, 2011 for a fuller analysis of these terms):

- *Diegetic.* Sound or music that is directly related, or at least perceived to be related, to the environment in which the subject is intended to be immersed. For example, in a movie this could be the sound coming from a television that is in the room pictured on screen. Another example would be the voices of the characters on screen or the sound of a character firing a gun or driving a car. In a nutshell, the subject is able to reasonably identify the source of the sound given the surrounding virtual environment
- *Non-diegetic*. These sounds are generally presented to augment or complement the virtual environment but come from sources that the subject cannot identify in the current environment. To go back to the horror movie example again, consider the famous shower scene from Alfred Hitchcock's classic *Psycho* from 1960: the screeching, stabbing violin sounds as the character of Marion Crane is stabbed by Norman Bates (dressed as his mother). There is no reason for the watcher of *Psycho* to believe that there are a collection of violinists in the bathroom with Norman and Marion, rather the music is there to enhance the environment that is presented.

Emotion in Computer Games

Game players exhibit larger emotional investment in games than in many other forms of digital entertainment, primarily due to the interactive nature of the medium. Jansz (2006) argues that game players often emotionally immerse themselves in games to experience emotional reactions that cannot reasonably be stimulated in the real-world: a sandbox environment for emotional development and experience. This notion will probably be familiar to most readers, as many of us will have deliberately watched a scary movie to try and frighten ourselves and because we enjoy experiencing the sensations and physical responses of being frightened, provided we are within a controlled environment.

Freeman (2004) provides a list of reasons that support the activation of emotion in computer games, citing "art and money" (p. 1) as the principle drivers, although his work focuses mainly on the latter, such as competitive advantages for games development companies, rather than direct benefit to consumers and game players. Nevertheless, as Freeman advocates, this awareness in the industry of the need to integrate emotion further into computer gaming, is evidence of market demand and big business interest in this exciting field.

Emotion manifests itself in many ways and there is an identifiable physical symptom in the user. Whilst the studies discussed later concentrate on identifying physical emotional reaction, these have not always been directly linked to the player's physical interaction with the game. However, research by Sykes and Brown (2003) describes an initial study that deals with investigating not just emotional *response* or *reaction* in users but emotional *interaction* with a game.

Sykes and Brown also support the theory that emotional reaction and interaction represent significant potential in being able to adapt and manipulate gaming environments in response to the emotional and affective states of the user. Their investigation dealt with determining if the amount of pressure applied to the buttons of a computer game controller pad correlated with an increased level of difficulty in the game environment. A benefit of using this approach as opposed to galvanic skin response or heart rate monitoring is that those mechanisms can be altered by the environmental changes around the user whereas changes in pressure applied to the game controller are much more likely to have been caused by events occurring in the game. Their results indicated that players did indeed apply greater pressure to the game controller when a greater level of difficulty and concentration was required in the game. Although the study is preliminary and relatively small-scale, the authors' methods of analysis employ significance testing of the data collected.

Ravaja et al. (2005) conducted experiments that attempt to evidence the impact of computer game-play upon human emotions by employing an array of biometric measurements. This is based upon the generally held theory that emotion is expressed by humans in three forms: "subjective experience (e.g., feeling joyous), expressive behavior (e.g., smiling), and the physiological component (e.g., sympathetic arousal)" (p. 2). Taking this further, the authors make the point that the psychological connection between a player and a computer game exceeds pure emotion and touches cognition where players make assertions and links to the game: believing they are a super-hero or ninja warrior, for example. This work also highlights the issue that, until recently, research into emotional enjoyment and influence has focused upon non-interactive, mass media communication channels, such as television, film and radio.

The wide range of measurements used by Ravaja et al. is concise and, as the authors indicate, few other studies have employed such a wide range of metrics when investigating emotional connection with computer gameplay. The authors use electrocardiogram (ECG)/inter-beat intervals (IBI), facial electromyography (EMG) and skin conductance level (SCL) as measurements during their experiments. The experiments showed that reliable results are achieved across a range of subjects in response to significant events in a game scenario (such as success, failure, poor performance and so on). This work provides very strong evidence that subjects exhibit strong, identifiable physical reactions that are typical during emotional arousal when playing with computer games. It supports the argument, made in this chapter, that emotion, through physical disturbance, is a strong method for detecting emotional state and response when interacting with computer games. Broadly speaking, positive and negative game events correlated to positive and negative emotional reactions in players. However, one point of note from the study is that the intuitively expected emotional response was not always the one that was encountered in subjects.

One criticism of Ravaja et al.'s study is that, although a reasonable sample size was used (36 participants), the gender balance was almost 70% in favour of male participants. Whilst it can be argued that the gaming population is likely to be male in majority, the study could have reflected the situation more accurately. The paper does not attempt to account for this disparity or investigate whether a significant difference was present between the results of the male participants and female participants (see Nacke & Grimshaw, 2011 for indications of gender difference in response to game sound). Although beyond the scope of their paper, the work could have been much strengthened by performing some form of subjective response with subjects on their performance in the game scenarios, thus allowing a more valid conclusion by employing triangulation of quantitative and qualitative methods. This would complement the reliable results attained through their objective measurements.

There is no doubt that emotion plays a significant affective role in computer gaming and that it has the potential to be used both as a reactive and interactive device to stimulate users. The emotion elicited in gamers is a function of both the content of the game as well as the context in which the user is placed, further justifying the aims and underlying concept of this chapter: that these three traits are inextricably linked and that further understanding and utilising them must therefore lead to more intense, immersive, and interactive gaming. Conati (2002), for example, considers how probabilistic models can be employed to develop artificial intelligence systems that are able to predict emotional reactions to an array of content and contextual stimuli in education games, with the aim of keeping the player engaged with the game. But what of sound linked to emotion in games?

The Use of Emotional Sound in Games

Research by Ekman (2008) bridges the gap between traditional movies and modern computer games by explaining how sound is used to stimulate emotions in each of these media. Ekman enhances her discussions with summaries of some of the numerous theories in the portrayal of emotional involvement experienced through sound and music. Perhaps most importantly in her work, Ekman emphasises the difference between the role of sound in movies as opposed to computer games. Principally, this is that sound in movies is present to enhance the narrative and heighten the experience whereas, in computer games, sound must perform not only this function but also serve as a tool for interaction, often to the extent where the narrative element is sacrificed in favour of providing informational content. Ekman's work therefore suggests that incorporating diegetic and non-diegetic sounds into computer games significantly increases the level of complexity for the sound designer.

Kromand (2009) feels strongly that sound can be used to influence a game player's stress and awareness levels by incorporating suitable mixtures of diegetic and non-diegetic sound. He provides examples of several contemporary computer games that feature such affective sound. In particular, his work focuses on the popular *BioShock*, *F.E.A.R.* and *Silent Hill 2* titles. Kromand's work is an interesting starting point and introduction to the use of sound in games, especially in inducing more unpleasant sensations. He provides extensive discussion and illustrative examples and considers the concept of *trans-diegetic* sounds (Jørgensen, 2011) those which transcend the traditional barrier between diegetic and non-diegetic. Kromand concludes by proposing that mixtures of diegetic and non-diegetic sound can lead to confusion and uncertainty about the environment and actions around the game player. He hypothesises that this confusion is purposefully implemented in the game environment and that the uncertainty of events taking place adds to the emotional investiture of the player in the game.

Though not as up-to-date as other works concerning computer games and human emotion, a corresponding work, which also looks at methods of eliciting emotional state in computer gamers, comes from Johnstone (1996). The age of this paper alone demonstrates the importance and significance of the emotional link between computer games and game players. His study concerns the discernment of emotional arousal by speech sounds made by users during their interaction with a computer game. Part of the rationale behind his approach is hypothesised to be because the feedback equipment of today (heart rate monitors and skin conductance devices) was not so readily or cheaply available in 1996.

An interesting concept that is partially addressed by Johnstone is that spontaneous emotional speech sounds differ acoustically from those that are planned and considered. If this theory holds true, then it means that genuine emotional responses can be distinguished from planned responses. In effect, this is somewhat analogous to the use of voice stress analysis in lie detection scenarios. Johnstone indicates that this ability is also useful in a truly interactive manner, since it not only means that users or game player responses can be analysed to determine emotional valences, but also that synthesised speech, such as the voices of characters in games, could be manipulated in similar acoustic ways to provide more realistic and affective game environments and conjunctions. For diegetic sounds in particular, this presents a world of opportunity.

The results of Johnstone's initial study are promising though there are some methodological aspects of the research that would have benefited from tighter control. For example, subjects' spontaneous speech sounds were recorded and analysed but they were also required to answer subjective questions to provide speech samples. By the very nature of such an enquiry, the subjects would have been required to consider their response during which time the effects of spontaneity or the moment could well have been depleted. The results gained are not enough to fully support the idea of distinguishing spontaneous sounds from planned although there is evidence to suggest that this might be a logical progression in future. Nevertheless, the data collected shows promise in being able to determine notions of urgency and felt difficulty in the game environment from events that are associated with achieving the objectives of the game. Primarily this can be measured by changes in spectral energy levels, low frequency energy distribution, and shorter speech duration.

In more recent work, Livingstone and Brown (2005) present theories and results that support the use of auditory stimuli to provide dynamic and interactive gaming environments. Whilst their paper explores the use of musical changes and emotional reactions in a general sense, part of their work is also devoted to investigating the application in gaming. Their underlying concept is that musical changes in the game can trigger emotional reactions in game players in a more dynamic manner than is currently the norm.

Livingstone and Brown employ a rule-based analysis of symbolic musical content that relates to a fixed set of emotional responses. Their work demonstrates that by dynamically altering the musical characteristics of playing music, such as the tempo, mode, loudness, pitch, harmony and so forth, the user perceives different emotional intentions and contexts within the piece of music that is currently playing. Music, then, stimulating one of the five human senses, is capable of influencing emotional change within humans in a computer game environment.

The work of Parker and Heerema (2008) presents a useful overview of how sound is used in diegetic and non-diegetic forms within computer games. They argue that greater use should be made of sound in order to enhance the game environment and experience. A primary exemplar used by them, is that sound should also serve as a tool for input and interaction with the game, rather than being present purely to be heard. They reiterate that sound in games at present is reactive rather than interactive. However, in this chapter, we suggest that sound is simply a tool of the emotions and that it is player emotion that should be interactive, rather than reactive, in order to provide a new level of computer gaming experiences. We feel this can be strongly underpinned by the use of sound.

Parker and Heerema go on to describe audio gaming and provide a series of examples and discussions of scenarios where sound can be used as the primary interaction mechanism between the player and the computer game. These range from the player reacting to audio cues, providing the game with input using speech, or other sonic input, and by directly controlling sound and music in the game. Although concise and valid at representing the current state of play of sound in games, their work does not consider the affective nature of using sound in games. Emotion is triggered by sound and the two are intrinsically linked.

Recent work by Grimshaw, Lindley, and Nacke (2008) seeks to formalise the relationship between a subject's immersion in a game environment as a function of the auditory content. Grimshaw et al. employ a series of biometric techniques to provide insight into the human emotional and physiological response to the sonic actions and environment of a first-person shooter game. Their method employs a significant array of quantitative, physiological measurements that are correlated with subjective questioning. The deep complexity of human emotion and psychology is exposed in their work as a strong relationship between the results of these two investigative methods cannot be found. This deficiency is the subject of significant discussion by the authors and, unsurprisingly, it is suggested as an area for significant future investigation.

It is important to place an emphasis on this point: although broad hypotheses and empirical evidence show sound and music play a large part in stimulating emotional responses in human subjects, the quantification of these effects, especially objective measurement, is elusive. Subjective investigation has traditionally always been the *forté* of psychological and sociological researchers. It is for this reason that sound designers and scientists working in the field must have an awareness of these issues, especially the sound designer working in computer game and multimedia development. In short, emotion is highly difficult to measure in an absolute way. Bridging this gap must be done carefully and backed-up by considered research and investigation.

There is a wealth of literature relating to the emotional impact of games. Equally, there is an increasing amount of published work concerning audio games; the majority of literature still concerns itself with traditional, visually-focused games. As the reader may have noticed in this section of this book, there are few studies that have concerned themselves with using sound as the primary interactive method whilst also monitoring and responding to the emotional reactions of the game player. It is just this sort of scenario that the studies and ideas presented in this chapter aim to inspire, support and help stimulate.

Are Sound and Music Really Important in Games?

It is interesting to consider to what extent sound is perceived as being important by users in computer games. If we consider the move from the beeps and clicks that early computer games such as *Space Invaders* and *Pong* made to modern alternatives such as the *Guitar Hero* series, we can see that the computer games industry has certainly placed an increased focus on the use of sound and music in games. To this extent, we conducted research, by means of a user survey, into determining user awareness of sound in computer games. The work is documented in grater detail in (Cunningham, Grout, & Hebblewhite, 2006), but a summary of the important findings and discussions are provided here.

Table 1. Overall Game Genre Preference of Survey Participants in Rank Order

Game Genre	Preference (%)
Role-Playing Game (RPG)	39
Shoot-em-up	24
Strategy/Puzzles	12
Adventure	9
Sports	9
Simulation	3
Other	3

This survey was undertaken to establish the various factors that subjects considered important when it came to purchasing a new computer game. Our initial hypothesis was that users would rate factors such as playability and visuals of a game, much higher than the sound and music, demonstrating that the focus in the computer gameworld tends to be in the areas of the graphical and gameplay domains. The survey had a total of 34 respondents. A profile of the gamers participating in the survey, in terms of their game type preference, is shown in Table 1.

We believe a future study should investigate whether the favoured game genre affects particular factors that users specifically look for in games. For example, role-playing games have been traditionally much more limited in terms of their graphic and aural flamboyance, with greater emphasis being placed upon story-line and depth, whilst action and adventure games are often much more visually stimulating.

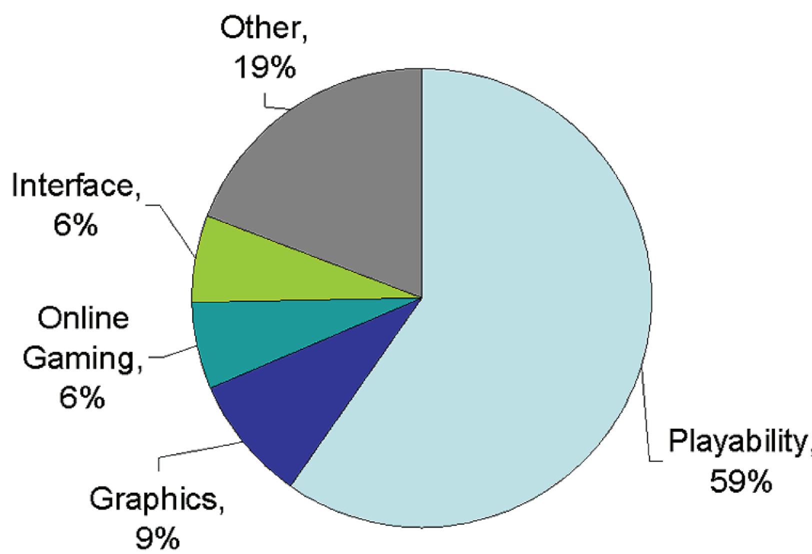
Figure 2 and Figure 3 illustrate the results of questions where participants were asked to indicate the most important and, since it was assumed prior to the study that the playability or gameplay would most likely be rated highly, the second most important feature that influenced game purchasing decisions.

Not surprisingly, we found that the most important factor is playability. The rating for all the other possible factors are negligible, although somewhat surprising is that no participants rated the sound or musical elements to be important when deciding upon a game to buy. Intriguingly, the ability to play a game online with other users took favour over sound, which is an intriguing insight into the mind of the 21st Century games player. Users who selected the "Other" category were prompted to provide a more detailed explanation. The responses received here all related to one of the following comments: "depth and creativity", "the whole package" and two participants stated that the "story or scenario" was most important.

It is argued, on the basis that playability will always rate highest, that the results in Figure 3 are more insightful than those in Figure 2. After all, the whole notion of computer games is that they are to be played with! This time we see, as we might well have expected, that the graphics and visual stimulation was the most popular factor. As expected, the sound present in a game was cited by a low percentage of those surveyed. The users who chose the "Other" category on this occasion also stated that the factor important to them related to the story of the game. Encouragingly, however, and still applicable in the context of sound in games, is the percentage of users that value the interface.

If we consider some of the most recent successful games, where the use of music and sound has been prominent, these titles almost all employ an interactive sound interface of some form. Prime examples include the *Guitar Hero*, *Rock Band*,

Figure 2. Rating the Most Important Game Feature



Dance Dance Revolution (DDR), and *SingStar* series of games, as well as *Battle of the Bands*, *Ultimate Band* and *Wii Music*, to name just a few. For the budding entrepreneur game developer, it is probably worth taking note that the majority of these titles revolve around the player being placed in a live music performance scenario or band. We briefly attempted to analyse these two assumptions through our survey, though the results are inconclusive with an almost 50/50 split between positive and negative responses. However, it is worth bearing in mind that these responses are now somewhat dated. An overview of the responses is presented in Table 2.

It is reasonable to suggest that the soundtrack of a game brings an added attraction when it comes

Figure 3. Rating the Second Most Important Game Feature

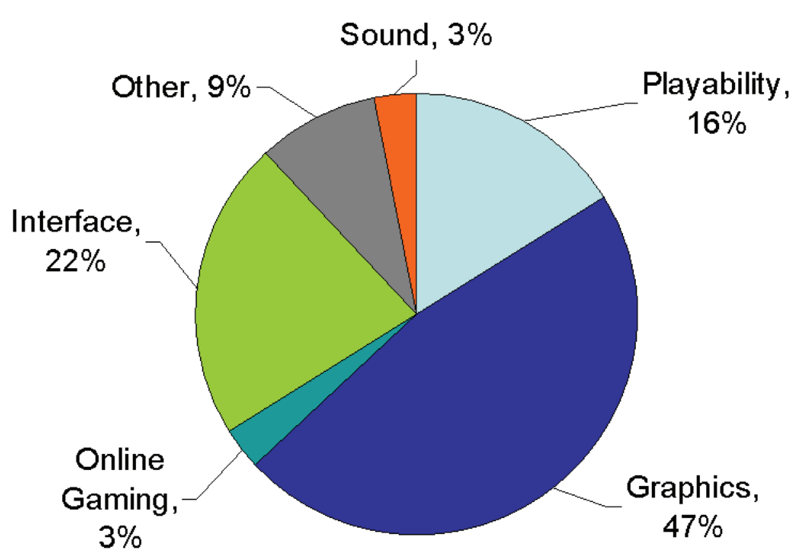


Table 2. Survey of Participants' Interest in Game Soundtrack and Music

Does the soundtrack/music of a computer game make you more interested in playing or buying it?	Yes	50%
	No	44%
	Don't Know	6%

to a gamer parting with their hard-earned cash. As mentioned earlier, game series like *Grand Theft Auto*, *FIFA* and *Dave Mirra* feature music by well-known recording artists, in some cases including music that has been commissioned specifically for that game.

It can be seen in the results summarised in this section that, other than the added value suggested above, users do not place any particular emphasis on game sound. As was expected, the main aspects users were interested in were the playability and graphics of a game, although interaction with sound offers great potential. The development of new sound-motivated games will be a dynamic and challenging field in the years to come, though we must not forget the golden rule of a successful game: playability.

The under-use of sound in games is further supported by Parker and Heerema (2008), a source the reader is encouraged to investigate if they are still in doubt as to the true potential of sound in the gaming environment. To quote directly from their work: "The use of sound in an interactive media environment has not been advanced, as a technology, as far as graphics or artificial intelligence" (p. 1). Their work goes on to justify these assertions and they explain that poor quality sound in a game often results in the game being unsuccessful in the marketplace, whilst the success of a game containing an acceptable or higher quality of sound will be based upon other factors such as playability or graphics.

It is clear from the discussions and investigation covered in this section that human interaction and psychological and emotional links with games are more and more to the fore, as well as becoming increasingly important in the development of successful gaming experiences. It is fair to assume that users are not only affected by sound and music but that they also respond to feedback and interact with the game, essentially providing full-duplex communication between human and machine that is becoming increasingly information-rich. It is these interactions and information that the rest of this chapter focuses on.

CONTENT

Digital audio data holds much more information than the raw binary data from which it is constituted. At its barest, sound and music are generally provided to augment and provide realism to the current scenario. However, as we demonstrated in the previous section of this chapter, computer games are truly multimedia experiences that combine a range of stimuli to interact with the user. In short, we see the area of content analysis as providing a semi-intelligent mechanism with which to tie together one or more media employed in a multimedia environment in order to provide even more effective and efficient interaction and experience.

Content of a particular medium can take many different forms, some of which will be shared across a range of media while others will be exclusive to a particular media type. The following is an attempt to briefly describe and exemplify these two categories:

- *Shared content information.* If we consider an entire multimedia artefact as being a hierarchical object, greater than the sum of its parts, then shared content information would be found attached to each of the media components present. For example, the publishing house, year of production, copyright information, and name of the game in which a multimedia object (a sound or otherwise) appears will always be the same. This information is generally that which is exclusively available in the form of meta-data and requires little data mining to extract

- *Exclusive content information.* This is information about the content that can only be found in a given type of media. Although the same content information may appear in multiple instances of that media, it will generally be exclusive to that *type* of media. For example, if we consider the music present in a computer game, the exclusive content information would include the tempo, amplitude range, time signature, spectral representation, self-similarity measurement, and so on.

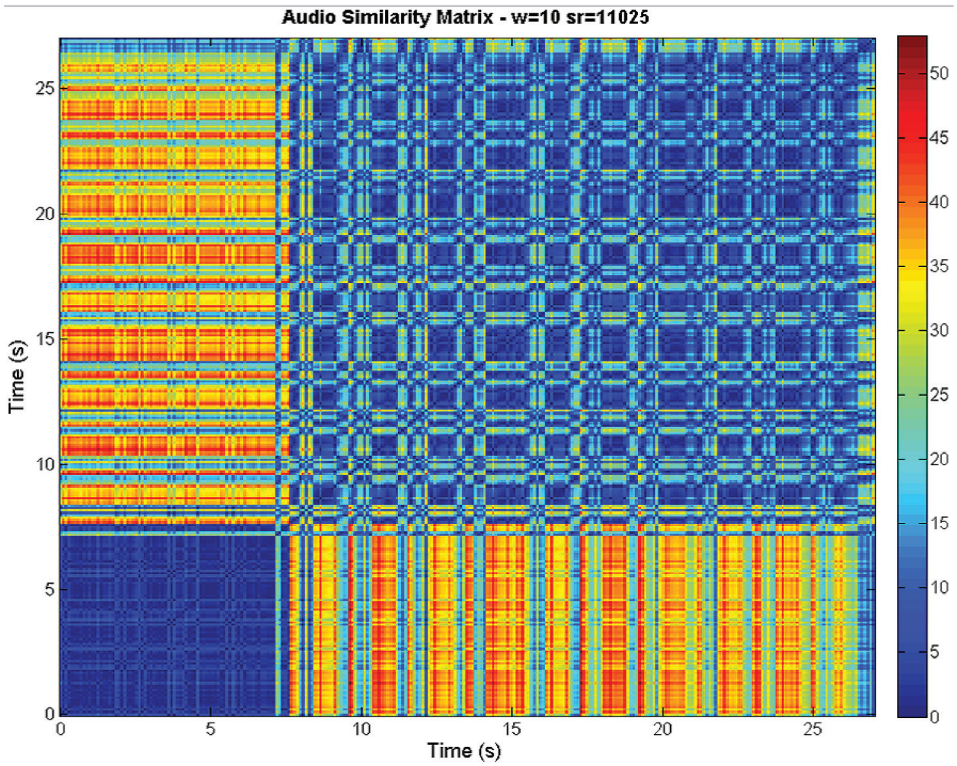
The relationship between sound and visual elements has been a mainstay of the media field since its inception. Consider the music video and Hollywood movie. Careful correlation occurs in these areas between the content presented to the user in these fields. Prime examples of these include the synchronisation between actions and transitions appearing in the visual field and the sound content. An illustration of this that the authors find particularly effective is in the opening sequence of the 1977 movie *Saturday Night Fever*. This particular scene sees the watcher treated to shots of John Travolta's feet, pounding the streets of New York in time to the Bee Gee's classic *Stayin' Alive*: a classic in its own right and an almost ridiculously simple example of the sound content being combined in the production of the visual content to produce something that has a much greater impact than either of the two individual components.

As Zhang and Jay Kuo (2001) demonstrated, it is quite possible to extract and classify a range of different sound content types from multimedia data, especially the kind of mixes found in traditional entertainment like television shows and movies. Though their work is focused on the traditional media of multimedia communication, the computer game environment is simply a natural extension of this, with the major difference being the integration of an element of interactivity.

It is these principles that we hope content analysis allows us to build upon and utilise in the field of electronic media processing and development. In particular, we hope that game sound content can be analysed to provide an enhanced gaming experience. As a good starting point for consideration, we began to explore the relationship between visual information and music in electronic media, to provide an augmented experience when viewing the visual data. In another of our works (Davies, Cunningham, & Grout, 2007) we attempted to generate musical sequences based upon analysis of digital images: in that particular case, those of photographs and traditional works of art. The underlying thoughts and questions that motivated that research revolved around suggestions such as: *What would the Mona Lisa sound like?* We felt this would also provide additional information for people who were, for example, visually impaired, and it could be used to provide added description and emotional information relating to a particular still image. It became a logical ethos that the only way in which this could be achieved would be to analyse the content of the image, as it is this that contains the information and components required to relay the same information but in an alternative format.

A tool that we have found very effective in analysing musical content is that of the Audio Similarity Matrix (ASM), based upon ideas initially proposed and demonstrated by Foote (1999). This allows a visual indication of the self-similarity, and therefore structure, of a musical piece. We suggest further reading into Foote's work as a

Figure 4. ASM of ABBA's "Head Over Heels" (28 second sample)



good starting point to stimulate the imagination into how content analysis can provide highly useful information for a variety of scenarios. A graphical example of an ASM is presented in Figure 4 as an exemplar, where dark colours represent high similarity and bright colours show low similarity.

Whilst we do not limit the application of content analysis to computer games, we suggest a few examples of appropriate situations where it may be used. Simple examples relate to the link between visuals and sound. In a game where the scene is bright and full of strong, primary colours, it would be pertinent to include sound that reflects this notion: bright and strong in timbre. On the other hand, a dark, oppressive scene would require slower, darker music with a thinner and sharper timbre, inducing a different set of emotions. In today's dynamic computer games, where the user has an apparently boundless freedom to explore a virtual environment, the dynamic updating of sound content to match the visuals requires some form of content analysis. Even a simple parameter that defined the "colour" of the scene or presence of tagged objects nearby would suffice on a basic level. Another suggestion would be to manipulate the gameplay by the choice of music and sound prescribed by the user. For example, the same game scenarios and task may be undertaken at different speeds, levels of difficulty, and in different environments, based upon the choice of music the user makes. Consider the scenario where a user may decide to play dance music whilst interacting with a game, thereby instigating a bright, quantised environment with predictable, rhythmic, structured gameplay content. Whereas if they choose highly random, noisy, alternative noise-core they would be presented with chaotic, overwhelming game scenarios: in both cases, a reflection of the structure and content of the music that can be achieved only through detailed signal and structural content analysis of the audio data.

CONTEXT

Context awareness also provides opportunities for a heightened user experience with digital media systems, particularly those that hold large data sets, the content of which may only be relevant to a user in certain usage scenarios. We believe the incorporation of contextual information into digital devices provides a more tailored experience for users. Contextual information can be considered as an added extra in digital media systems, allowing more defined information about the user to be brought into software systems. Recommendation systems, for example, are a great example of where contextual data can be included.

Schmidt and Winterhalter's (2004) work in e-learning is a good example of how context awareness can be incorporated into digital, computer-based communication media. In their field, the context of the user is particularly important as it allows greater control and focusing of learning and teaching materials in order to engage at a deeper level with the user. Their work emphasises that the key stages of context awareness are in first acquiring contextual information and then building a suitable user-context model so as to estimate the current context of the user. Schmidt and Winterhalter also reinforce notions that good contextual modelling comes by acquiring information from a range of sources. Most importantly, in discussions of the importance of user context, Schmidt and Winterhalter hit upon the key questions that context awareness is able to begin to address: "How do we know what the user currently does, or what he intends to do?" (p. 42).

Schmidt and Winterhalter choose to employ more passive mechanisms for contextual data acquisition, such as those which passively track user progress through tasks and record commonly accessed information. This is perfectly suitable for e-learning applications but, in the field of computer games and interactive entertainment, we feel that something a bit more fortuitous is required.

A crucial work that backs up these notions of more interactive and reliable context awareness, especially when it comes to the surrounding environment, can be found in Clarkson, Mase, and Pentland (2000). Although this work may now be slightly dated, the principles and techniques employed in their work are effective and provide good examples of the type of contextual information that can be acquired by using simple sensor input. Their work investigates how context, such as whether the user is on a train or at work and whether they are in conversation or not, can be estimated from sensor input, primarily a camera and microphone. Such work provides a strong basis from which to lead into more specific analysis of context that is relevant to the current activity or software application. This is further elaborated upon in the context of mobile device usage by Tamminen, Oulasvirta, Toiskallio, and Kankainen (2004), who consider determining contextual information in mobile computing scenarios.

Computers, gaming consoles and mobile devices have all become much more powerful in recent years and interface with a range of local and remote information sources. These information sources range from the traditional tactile input devices to accelerometers, touch screens, cameras, microphones and so on. The Nintendo Wii and Apple iPhone and iPod are prime examples of such low-cost, sensor-rich, powerful computational devices. The technology available in these devices, as well as those devices that can be further added into the chain, mean a wide range of contextual information can potentially be extracted from a game player, be they mobile or static.

We consider that the foremost sources of contextual information come from the user themselves and from the surrounding environment in which the user is currently immersed. This is further ratified by Reynolds, Barry, Burke, and Coyle (2007) who also consider the importance and usage of contextual input parameters from these two domains in their own research.

Information from the user is arguably the most useful data that can be acquired in determining the context of the user. This allows the researcher to begin to investigate factors such as their level of activity, stress levels, emotional state, for example. We propose mechanisms such as skin conductivity, motion and heart rate data that might be acquired directly from the user and would prove particularly useful in monitoring their contextual state.

Factors in the environment around the user are likely to have an effect on their performance in a game, their general attitudes, and their emotional state. A number of metrics can provide suitable input to a software system to estimate environmental context. Environmental information includes the amount of ambient noise, light levels, time of day and year, temperature and so forth. We feel that the devices and information in this scenario are relevant to many contextual extraction applications, not only those of digital entertainment and games.

Hopefully, the reader can begin to gain an insight into the usefulness of contextual information from the examples and discussions in this section of the chapter. The next section of this chapter seeks to exemplify how context (as well as emotion and content) can be employed in digital multimedia applications, especially those that relate to sound. We feel that, in computer games, the virtual gameplay environment can be tailored to reflect the real environment of the player. In all, this will provide a deeper, more immersive experience: this will help the player to develop greater emotional and personal investment in the game. It will also be interesting to see if such a game can contribute to altering the emotional state of the user and impact upon their own personal context. For example, can games be designed that would relax a user, reduce their stress levels and heart rate, and even make them alter their surrounding environment to reflect their new, calmer state? Only through more contextual awareness and pervasive interactions with games will we know the answer to this question and others.

DETERMINING USER PERCEPTIONS OF MUSIC

In this section of the chapter, we aim to gain more of an insight into how emotion, content and context are attached to music by human listeners. By investigating the various perceptions and semantic terms users relate to different musical genres, it is possible to gain a deeper understanding of the ways in which humans relate their emotions, musical content and the context of different types of music.

Wide ranges of semantics are frequently employed to portray musical characteristics and range from technically-related terms to experiential narratives (Károlyi, 1999). It is proposed that the characteristics of a piece of music are difficult to quantify in a single term or statement. Whilst high-level abstractions may be possible that categorise the music or provide an overview of the timbre, this is a highly subjective and individual (and potentially emotionally influenced) expression of a listener's experience of the music. Such an investigation also allows groupings to be applied to terms, understanding to be formed and a mapping of the relationships between these groups to be formed.

Repertory Grid Technique

In order to extract common descriptive features and semantics that are most meaningful and globally understood, it is better to employ a technique where the listener subject may employ their own descriptions of the elements under investigation. George Kelly's work (1955) into *personal construct theory (PCT)* and *personal construct psychology (PCP)* provides a suitable mechanism, known as repertory grid analysis, by which such descriptions can be elicited from subjects, correlated and employed in measurement subject experiences. Kelly's work in this area is grounded in principles of *constructivism*, where subjects identify and deal with the world around them based upon their own experiences and interpretation of events and objects.

Repertory grid analysis consists of defining a particular *subject* or *domain* to be investigated within a particular *context*. Descriptions of instances or examples of the domain are known as *elements* and bi-polar descriptions of the elements, known as *constructs*, are rated on a scale (usually 5 or 7 point). For example, the domain being investigated might be movies and the elements could be a number of popular movies and the bi-polar constructs used to describe and rate the movie elements could be violent or non-violent, an adult's film, or a children's film and so on.

Constructs are defined by the subject with the help of the interviewer, who enables the subject to produce more constructs by defining the relationships and differences between the nominated elements. This can be enhanced through interview techniques involving triads, where three random elements are chosen and the subject asked to choose the least similar of the three and define the construct that separates it from the other two elements (Bannister & Mair, 1968). Subjects then provide a rating on the point scale for each element against each of the constructs they have defined in order to complete their grid. Alternatively, and particularly of use when subjects struggle to separate two elements from a third, the interviewer can also find it useful to present a subject with two elements and ask them to explain the factor that differentiates the two elements. This will often provide one pole of a construct and the subject is then asked for what the opposite side of that particular construct would be.

Once a desired number of subjects have completed a repertory grid each, the grids are then concatenated and can be immediately visualised as one large grid but, more crucially, the opportunity is available to determine the importance of elements and constructs within the larger grid.

The *bi-polar* nature of defining constructs allows the context and relationship of a construct to be articulated and better understood by the researcher. This further removes ambiguity when a subject provides a rating, since the interrelation between the opposing ends of the scale have been specified by the subject themselves (Kelly, 1955). Further detail of PCP and repertory grid technique goes well beyond this work and can be found in Kelly's seminal text.

Using a Repertory Grid to Understand Perceptions of Music

A set of elements was defined to include a representative spectrum of musical genres upon which the subjects would be asked to define their own bi-polar personal constructs in regard to their experiences and perceptions of the characteristics of those genres, in their experiences of listening to music. Whilst there are many sub-genres and pseudo-related musical styles, this provides an appropriate, broad spread for the purposes of this particular investigation without making the interview process for the participant overly laborious in terms of time and effort. The elements defined were those shown in Table 3.

Subjects for the investigation were drawn from a random sample of the population. Subjects were interviewed on an individual basis and told that the purpose of the exercise was to get them to express their perceptions about the characterising features of different type of music. To carry out

Table 3. Musical Genre Elements Used in Repertory Grid Experiment

• Pop	• Rock	• Dance
• Jazz	• Classical	• Soul
• Blues	• Rap	• Country

the rating of elements against their defined constructs, subjects were asked to perform a card sorting exercise for each pair of constructs. The use of *triads* was made to elicit the choice of constructs by randomly selecting 3 elements and once subjects began to struggle with the use of triads they were asked to differentiate between 2 randomly selected elements. A total of 10 subjects were selected to participate in the elicitation process of the repertory grid interview. The age of subjects interviewed ranged from 16 to 59, with the average age being 34, and there was a 50/50 male/female gender split. The results of the ten repertory grid interviews are presented in Figure 5 and Figure 6.

Though the number of subjects involved in the repertory grid interviews appears to be a low population sample at first glance, the granularity from these interviews comes from the sum number of constructs elicited across all participants. Furthermore, the data retrieved using constructs provides both qualitative and quantitative information regarding the domain of enquiry.

In addition to the visual analysis of a repertory grid, a PrinCom map, which makes use of Principal Component Analysis (PCA), can be derived that relates elements and constructs in a graphical fashion where the visual distance between elements and constructs is significant. The PrinCom mapping integrates both elements and constructs on a visual grid and shows the relationship between the two. A PrinGrid for the repertory grid derived in this investigation is shown in Figure 7.

It is the *constructs* elicited that are particularly of interest within the scope of this work. The constructs used by subjects provide insight into how they perceive music. As can be seen from the grid in Figure 5 and Figure 6, the range of constructs elicited provides an insight into, not only how subjects typically perceive the sound content of each musical genre but also, terms relating to the context in which they place each genre and occasional indications of the emotional impact of each genre. For example, by also looking at Figure 7 we can produce the notion that blues music is *"*emotionally evocative*"*, has "specific geography & history", is placed in the context of being "African American", and is "mellow". Naturally, there is some subjectivity present here and these statements are open to interpretation, but to most readers it is expected that these constructs should represent the group norm.

A perceived limitation of the repertory grid technique to have been encountered during this particular study is that of familiarity with the elements under investigation by subjects, during repertory grid interview. During interviews there were clearly some elements that subjects were definitely not as familiar with as others. It was observed that subjects would often group together the elements they were less familiar with when rating elements against their chosen constructs. Whilst it is appreciated that this phenomenon is likely to be particularly present in this study, due to music awareness firmly depending upon personal preference or taste, it is doubtless likely to occur in other scenarios.

Using a repertory grid sought to elicit human-friendly descriptions of musical characteristics. Although not strictly timbral definitions, these constructs succeed in describing the characteristics of musical genres. To put this into the context of artistic definitions with the notion of a visual metaphor, whilst *timbre* is a human description of the *colour* of a sound or piece of music, these *constructs* can be thought of as describing the *patterns*; the mix of shapes and colours that provides deeper information about the content and the bigger picture

We find repertory grid investigation to be a highly useful tool in determining group norms and perceptions of important factors in any field that is being explored. In the context of this chapter, it can hopefully be seen that using such techniques would allow information about how a group of users would perceive a game and game sound in terms of the content that constitutes the game along with their emotional perceptions of the game and the context in which they view it.

Figure 5. Repertory Grid Ratings of Musical Genres (Part a).

Display
"Identify common artefacts and descriptions of musical genres"

	Pop	Rock	Dance	Jazz	Classical	Soul	Blues	Rap	Country	
Mellow	3	5	3	1	2	1	1	5	2	Heavy
Melodic	3	5	4	1	2	1	1	5	3	Less Melodic
Serious	5	1	5	1	1	2	1	4	3	Less Serious
Old	5	3	5	1	1	2	1	4	2	New
Large number musicians	5	4	5	1	1	4	4	5	5	Small number musicians
Commercial	1	3	1	4	4	4	5	3	4	Less Commercial
Less Diverse	2	2	3	4	3	3	3	2	2	Diverse
Lyrical Expression	3	5	3	5	5	3	3	1	2	Musical Expression
Serious	5	3	5	1	1	2	2	4	2	Pretentious
Unstructured	5	4	5	2	5	5	3	5	5	Structured
Non-commercial	5	4	5	2	1	4	2	5	3	Commercial
Non-narrative	2	3	1	1	3	4	4	5	5	Narrative
Un-resolving	3	5	1	2	5	4	5	4	5	Resolving
Uninteresting	2	4	1	4	4	3	4	3	2	Interesting
Weak Rhythm	4	4	5	1	3	3	3	5	1	Strong Rhythm
Emotionally Neutral	2	4	1	1	4	3	4	2	2	Emotionally Evocative
Narrow Genre	2	1	2	3	5	2	2	1	3	Broad Genre
Freeform	5	5	5	3	3	4	4	4	4	Formulaic
Corporate	1	3	3	3	1	5	5	4	5	Communal
Intentionally offensive	5	1	2	3	3	3	3	1	5	Intentionally bland
Inaccessible	5	4	3	1	3	3	4	2	4	Accessible
African American	5	5	3	3	5	2	1	1	4	White American
Diverse Geography & History	3	2	3	4	4	4	5	5	4	Specific Geography & History
Young people	2	2	1	5	4	3	3	1	4	Old people
Commercial	1	1	1	4	5	3	3	2	4	Non-commercial
Dancing Music	1	3	1	4	5	3	4	2	4	Listening Music
Instrumental	3	3	3	2	1	2	2	5	5	Vocal
Electric	1	1	1	5	5	3	5	1	5	Acoustic
Large Scale	5	5	5	3	1	5	3	5	5	Small Scale
Melodic	1	3	5	3	3	3	3	4	2	Rhythmic
Playing Instruments	5	1	3	1	1	3	1	5	1	Not Playing Instruments
Relaxed	2	3	5	1	1	1	3	5	1	Un-relaxed

EXAMPLE APPLICATIONS IN CURRENT RESEARCH

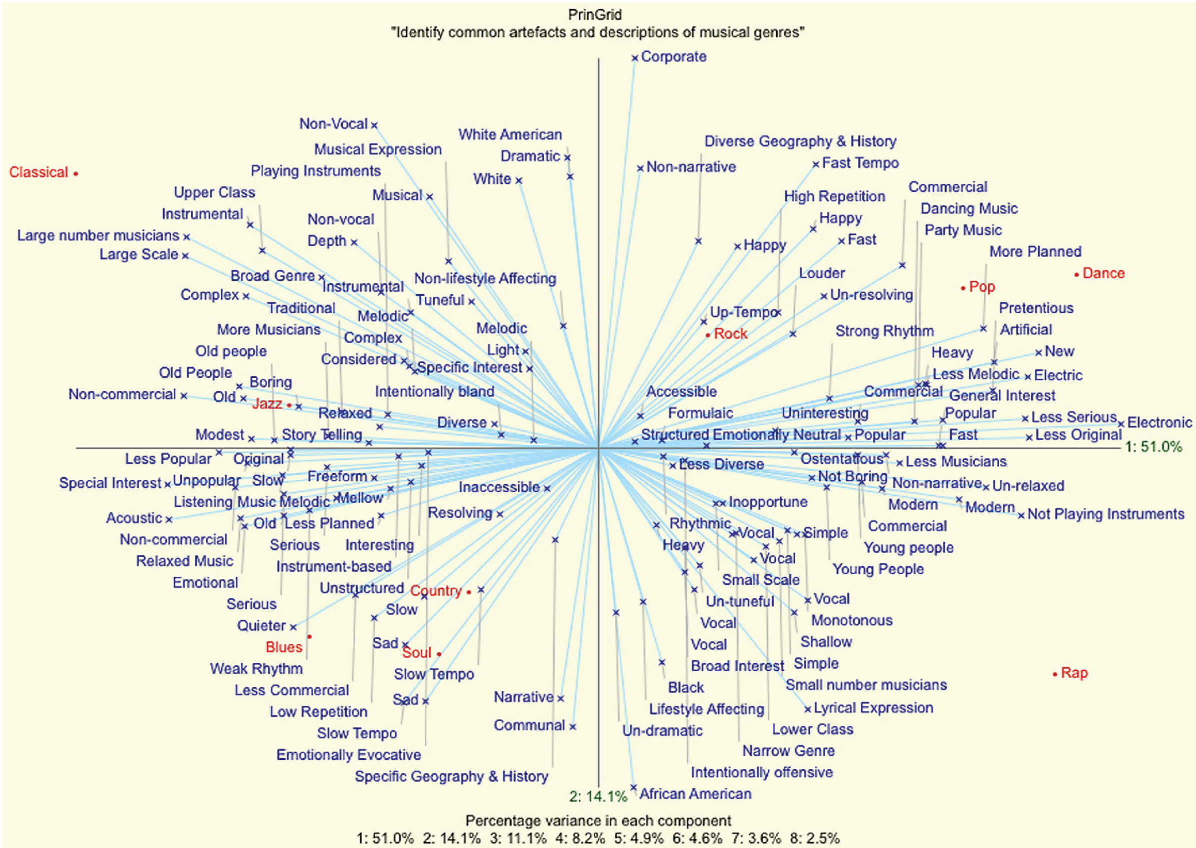
Presented in this section are summaries from some research work that has been influenced or involved by the use of emotion, content and context in various guises. A number of the works presented here have been studies involving a small set of music. For convenience, this small database of music is shown in Table 4 so that the reader may

Figure 6. Repertory Grid Ratings of Musical Genres (Part b).

Display
"Identify common artefacts and descriptions of musical genres"

	Pop	Rock	Dance	Jazz	Classical	Soul	Blues	Rap	Country	
Vocal	1	2	3	3	5	1	2	1	1	Non-Vocal
Less Original	1	5	1	4	5	3	5	1	5	Original
Slow Tempo	5	4	5	2	3	1	1	2	3	Fast Tempo
Traditional	3	2	5	2	1	3	1	5	2	Modern
Popular	1	1	2	5	5	4	4	1	3	Less Popular
Quieter	5	5	5	2	3	3	2	4	2	Louder
Instrumental	5	5	4	2	1	4	4	5	5	Vocal
Slow Tempo	3	4	5	3	3	1	3	3	3	Up-Tempo
Instrument-based	3	1	5	1	1	1	1	5	1	Electronic
Tuneful	2	3	4	3	2	3	3	5	4	Un-tuneful
Complex	3	2	5	3	1	3	3	4	3	Simple
Slow	3	4	5	1	3	1	1	3	3	Fast
Un-dramatic	3	4	3	2	4	2	2	1	1	Dramatic
Low Repetition	4	4	5	3	3	1	2	4	3	High Repetition
Non-vocal	4	4	3	3	1	3	3	5	4	Vocal
Sad	5	4	4	2	3	1	1	3	3	Happy
Boring	4	5	4	3	1	2	3	5	3	Not Boring
Slow	4	3	5	2	1	1	3	5	2	Fast
Special Interest	5	5	5	1	1	1	3	5	3	General Interest
More Planned	2	3	1	5	5	5	4	3	4	Less Planned
Less Musicians	3	2	1	5	5	5	4	2	1	More Musicians
Old People	5	3	5	3	1	3	3	5	3	Young People
Shallow	1	4	3	4	5	1	4	1	2	Depth
Artificial	1	3	1	3	5	4	5	1	5	Emotional
Relaxed Music	5	3	5	4	1	2	1	5	2	Party Music
Unpopular	5	3	5	2	2	3	2	5	3	Popular
Specific Interest	5	1	1	4	2	2	3	5	2	Broad Interest
Sad	5	3	5	5	3	2	1	3	2	Happy
Inopportune	1	4	2	1	5	2	4	2	2	Considered
Ostentatious	1	1	2	2	5	3	4	1	3	Modest
Lifestyle Affecting	5	2	4	3	4	4	3	1	3	Non-lifestyle Affecting
Modern	2	2	1	4	5	2	4	1	4	Old
Vocal	2	1	4	3	5	2	3	1	1	Musical
Simple	2	2	1	4	5	2	3	1	2	Complex
Monotonous	4	4	1	2	5	4	3	1	4	Melodic
Light	1	5	4	3	2	3	3	4	4	Heavy
Non-narrative	3	4	1	2	5	4	4	1	5	Story Telling
Lower Class	2	2	1	4	5	2	2	1	2	Upper Class
White	3	2	3	4	1	5	5	5	1	Black

Figure 7. Musical Genre Principal Components Analysis.



cross-reference the ID number to the song, where appropriate. We feel that this small selection of songs represents a reasonable cross-section of contemporary popular musical genres.

Responsive Automated Music Playlists

Some of our most recent and cognate work combining the use of emotion, content and context in musical applications has been in the area of intelligent playlist generation tools and this work is explored in greater detail in a separate work (Cunningham, Caulder, & Grout, 2008). However, to see the effectiveness of combining all three of these areas, the reader is provided here with a summary of that work to date.

Our main motivation in this area of research and development was to address some of the shortcomings traditionally employed in automatic recommendation and playlist generation tools. Historically, these tools evolved in a similar way to that of Automated Collaborative Filters (ACFs). That is to say, simple measurements of user preference and the preference of a typical population were used to build a ranked table of music in a library. These analysed information such as the most frequently played tracks, a user rating of each track, favourite artists and musical genres, and other meta-data attached to a song (Cunningham, Bergen, & Grout, 2006). However, this is not to totally trivialise the area of automatic playlist generation, since a number of systems exist that employ much more advanced learning and

Table 4. Mini Music Database Used in Testing

ID	Artist	Song
0	Daft Punk	*One More Time (Radio Edit)*
1	Fun Lovin' Criminals	*Love Unlimited*
2	Hot Chip	*Over and Over*
3	Metallica	*Harvester of Sorrow*
4	Pink Floyd	*Comfortably Numb*
5	Sugababes	*Push The Button*
6	The Prodigy	*Breathe*
7	ZZ Top	*Gimme All Your Lovin'*

analysis techniques and technologies (Aucouturier & Pachet, 2002; Platt, Burges, Swenson, Weare, & Zheng, 2002).

In recent years, as computational power and resources have increased, the tools that underpin musical and sound content analysis have migrated deeper into the field of playlist generation, allowing greater scope and accuracy for classification of musical features (Logan, 2002, Gasser, Pampalk, & Tomitsch, 2007). However, although these advances have been significant, such methods have always focused on musical and sonic information extraction and few systems have considered the wider scope of the user and his or her environment. It is reasonable to expect that factors relating to the emotion, state of mind and current activities of a listener will greatly influence their current and, most importantly, next selection of music.

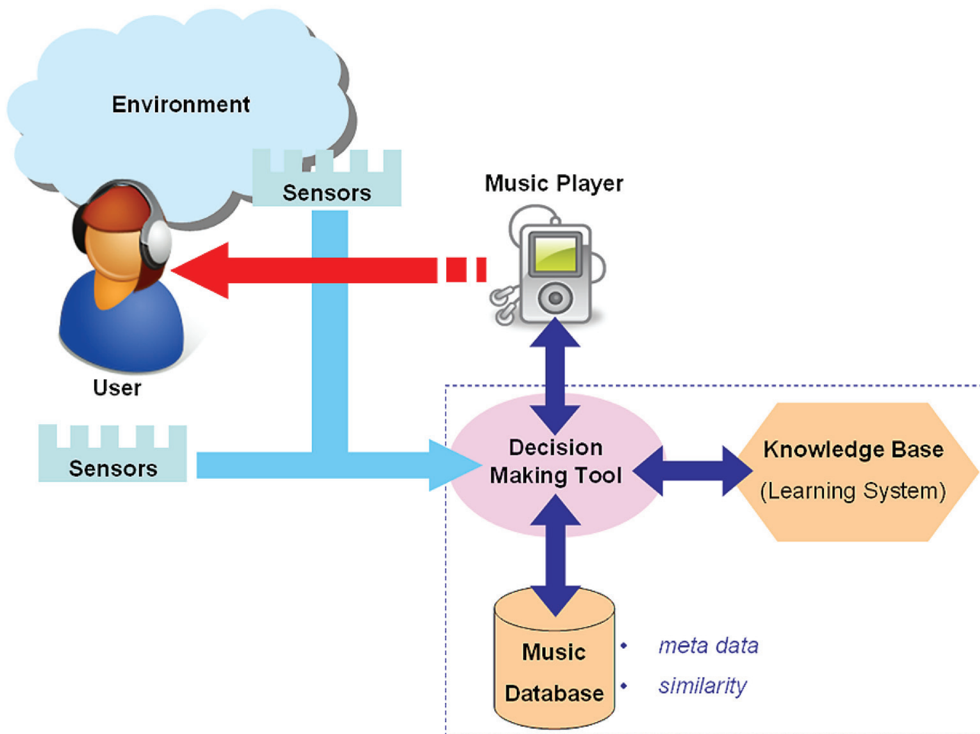
Emotion, Content and Context in Automatic Playlist Generation

In this review of our recent work in the area of intelligent automatic playlist generation, we provide details of the development principles, investigation and analysis into the viability of playlist generation that considers the wider circumstance of the listener. To achieve more accurate and useful playlist generation, we propose to not only build upon the established principles of using information about the musical *content* but also to examine the *context* the listener is in and how these external factors might affect their choice of suitable music. Additionally, the emotional state of the user is of interest since this also is likely to influence their choice of music. To visually summarise the complete system we are describing here, a figure showing an idealised scenario is provided in Figure 8. To realise a system that will potentially have to deal with and correlate a wide range of input parameters, we employ approaches that utilise fuzzy logic and self-learning systems.

Principally, determining the user's emotional state is of key importance to a truly successful and useful playlist generation system. This is informed by and correlated with the state of the surrounding environmental factors for the user, as well as their current levels of movement, physical activity, heart rate, stress levels, and so on.

Within our work, we felt that it was initially most important to investigate the current locomotive state of the user. For example, a user who is moving a lot and accelerating rapidly may be participating in an energetic activity such as running, dancing, cycling, or exercising in some way. It is feasible to suggest that most people listening to music in such scenarios would be likely to want to listen to music that reflects the nature of their physical motion such as music with a dominant, driving rhythm and high tempo, greater than 120 or 130 beats-per minute, for instance.

Figure 8. Emotion, Content and Context Aware Playlist System



Given currently available technology, it is also relatively easy to find equipment that allows the measurement of a user's movement. This was realised in our case by employing the wireless hand controller from the Nintendo Wii: the Wiimote. The Wiimote, when compared to other alternatives, is a cheap device that allows measurement of three-dimensions of movement. The Wiimote is almost universally accessible since it employs the Bluetooth communication protocol to send and receive data to a paired host. As Maurizio and Samuele (2007) demonstrate, valuable motion information can be retrieved through the accelerometers contained in the Wii controller.

Implementation and Initial Results

Our initial work in this field sought to demonstrate the ability to attain, analyse and correlate content-related data about music and contextual information acquired from the user to arrive at an estimate of the user's emotional state (*E*-state). To achieve this, we developed a small scale system that would work from a number of simulated factors (controlled by the researcher) and also live data extracted from sensors, principally the Wii controller. This system was designed to work with a small music database consisting of eight songs, shown in Table 4, and rank these songs in order of most suitable, based upon the estimated E-state.

To begin working with the motion data from the Wiimote controller, we attempted to work with four simple locomotive states: standing, walking, jogging, and running. These simple locomotive states were believed to be detectable from not only the Wiimote but a range of motion measurement devices such as the accelerometers built into the Apple iPhone/iPod, as well as higher-level systems such as a Qualisys motion capture system (which we also had access to and allows us to verify the

Table 5. Defined Set of Emotional States (E-states)

E-State	Depressed	Unhappy	Neutral	Happy	Zoned
Numeric Range	0-3	3-4	4-6	5-8	7-9

results obtained from the Wiimote). Similarly, for each of the other input parameters we work with, such as weather conditions, amount of light and so on, a range of states was also defined. As previously mentioned, due to our current limitations of time and resources, we focused on only the locomotive state of the user. It is hoped that in future, further ratification will be achieved by using other user measurements such as heart rate and galvanic skin response.

To avoid additional complication, and allow greater control over the testing procedure, this set of parameter ranges and values was loosely defined according to the empirical and historic knowledge of the individual. However, this is too fixed and logical to fit the way things tend to work in the real world: therefore, to make these values more representative of real scenarios, they are *fuzzified*, when defined in the fuzzy logic system. In simple terms, this means the boundaries and degree of accuracy of each point on a scale is related within the range of all available values.

The implemented fuzzy logic system provides a single output value that represents the predicted *E*-state of the listener. For simplification, we began by defining five emotional states and assigned a numeric range to each state to allow the representation of varying degrees of this state and the overlap where states merge into one another. This is appropriate, since it is very difficult to place an absolute, quantitative metric onto the complex emotions felt by humans. A table representing these allocations is shown in Table 5.

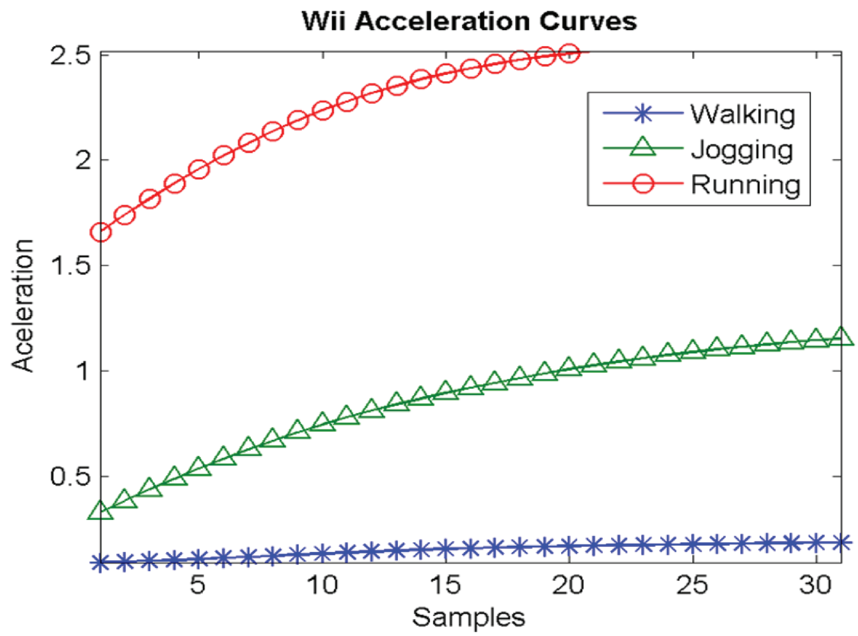
To verify the effectiveness of using the Wii controller as a device to measure movement, and particularly locomotive state, we performed a number of experiments benchmarked against a Qualisys motion detection system. By determining the rate of acceleration from the accelerometers in the Wii controller and asking a subject to provide three locomotive states (walking, jogging, and running) we can see that the Wii controller allows rapid identification of each of these states, as the graph in Figure 9 shows.

We defined a number of different scenarios, that combined a range of contextual parameters which a user might typically find himself or herself in. Each of these scenarios is shown in Table 6.

These parameters included those that the user has control over, in this case the locomotive state, and external, environmental factors, beyond the control of the user. We carried out a quantitative investigation, using 10 subjects, where each subject was asked to map one of the emotional states from Table 5 against one of the scenarios from Table 6. From this investigation, the average *E*-state response rating for each scenario is shown in Table 7. Although the use of an average response is not ideal, it provides a sufficient insight into the common perception of each scenario and when we performed analysis of each response, there was strong correlation across all of the subjects.

Each of the songs in the mini-database was allocated a range of values from the *E*-state table by the research team. These values reflected the researchers' perceptions of the content and emotional indicators present in the music. Naturally, in future research, we will explore the perceived emotional state attached to each song, by employing a more detailed sample of a suitable population. However, for now, these allocations were decided to be a controlled factor in this particular investigation. These allocations were then mapped

Figure 9. Wii Acceleration Curves for 3 States of Locomotion



against the emotional states extracted from the user-scenario study and each song's *E*-state was ranked against each scenario's *E*-state to provide a grade for each song, using a simple Euclidean distance measurement of the form

$$G(p,q) = \sqrt{(p-q)^2}. \quad (1)$$

Table 8 shows the resultant ranking of songs (by their ID), for each of these scenarios.

At this stage, it is recognised that the system is currently more limited than the idealised scenario presented earlier in Figure 8. A number of factors have been simulated and a number of assumptions have been made. However, using a Fuzzy Rule Based System (FRBS) we have been able to implement a limited version of the system outlined in that dynamically outputs an E-state based on live sensor data from a Wii controller and simulated environmental parameters. An outline of the Takagi-Sugeno-Kang (TSK) type

Table 6. Range of Scenarios in Subject E-state Evaluation

ID	Scenario Description
1	Walking, temperature is hot, lighting is dark/grey and weather is light rain.
2	Stationary, temperature is cold, lighting is dark and weather is raining.
3	Stationary, temperature is warm, lighting is brightening and weather is dry.
4	Running, temperature is hot, lighting is daylight/getting brighter and dry.
5	Walking, temperature is getting hot, lighting is dark and weather is drizzling.
6	Stationary, temperature is hot, lighting is grey and weather is dry.
7	Walking/Jogging, temperature is mild, daylight and it's dry.

Table 7. Average User E-state for each Scenario

Scenario ID	Emotion (0 = Unhappy; 100 = Very Happy)
1	4.3
2	0.0
3	6.8
4	7.7
5	3.0
6	3.8
7	6.5

FRBS we employed, along with the *fuzzified* input parameters, is shown in Figure 10.

Whilst a number of parameters such as light and temperature measurement have been simulated at this stage, the implementation of live sensor data from such devices is a trivial one and is only limited by the current lack of the hardware resources to incorporate these devices into the live system.

In its current state the system demonstrates the ability to read contextual data from the user and correlate this with information from the environment to make an informed judgement of the user's emotional state. With future development and also feedback from the user whilst using the system, the facility will be available to teach the playlist generator about the user's preferences and train the accuracy with which the system is able to estimate the emotional state of the user.

FUTURE RESEARCH IDEAS AND CONCLUSIONS

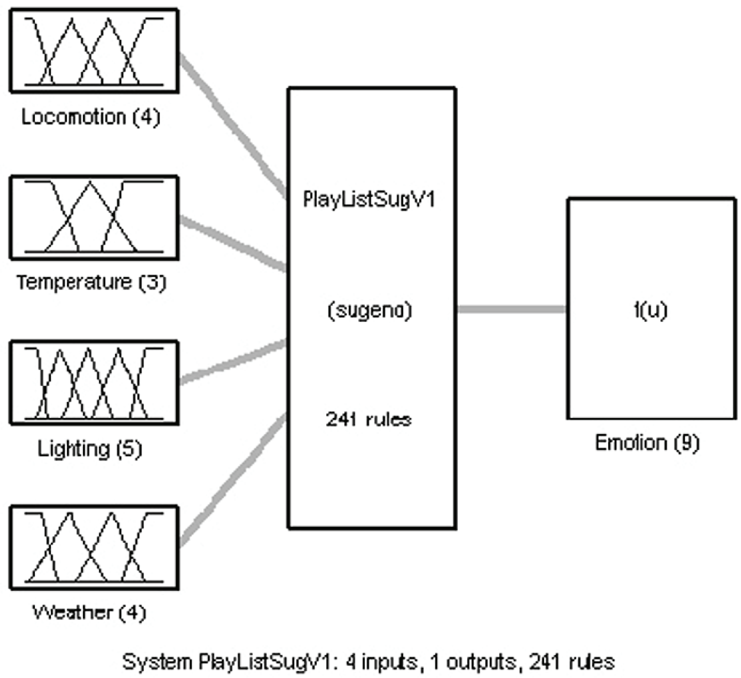
We have seen that awareness of the presented issues is beneficial in not only providing richer interactive experiences and more appropriate information, but that knowledge of the purpose of information can be used to optimise computational challenges. Furthermore, information, such as that presented visually, can be analysed in terms of content and context with the goal of being able to stimulate emotions in a user who might otherwise be disadvantaged from such an experience, due to visual impairment.

It is hoped that we have demonstrated the applicability of determining and analysing features related to emotion, content and context in relation to improving systems where user-interaction is of particular significance.

Table 8. Ranked Playlist Ordering for Set of Given Scenarios

Scenario	*E*-state	Playlist order
1	4.3	1; 4; 0; 7; 3; 6; 2; 5
2	0	6; 3; 4; 1; 0; 7; 2; 5
3	6.8	0; 7; 2; 5; 1; 4; 3; 6
4	7.7	2; 5; 0; 7; 1; 4; 3; 6
5	3	4; 3; 6; 1; 0; 7; 2; 5
6	3.8	4; 1; 3; 6; 0; 7; 2; 5
7	6.5	0; 7; 1; 2; 5; 4; 3; 6

Figure 10. Simplified Overview of TSK-type FRBS Used in Playlist Generation



For the budding researcher interested in these areas, we suggest the following broad, non-exhaustive, list of some of the key research themes and areas that would greatly benefit from further investigation:

- Explore the commonly perceived emotions of users playing computer games and determine the most reliable methods with which to record and model these emotions
- Develop a common software toolbox to allow audio content analysis to be easily bolted into a range of software products
- Further examine the value of environmental context for users playing computer games compared with user-centred contextual data
- Assess gameplay parameters that are best influenced and reflected in the emotion, content and context of the user
- Develop fuzzy logic systems that can accurately read a range of content and contextual data and output a robust, truly reflective emotional state for the majority of a sample user population.

Above all, we hope that in reading this chapter we have stimulated intellectual thought and got the creative juices flowing. Our aim here is not to provide a cast-in-stone set of data and instructions for the budding developer and researcher to follow and obey: far from it! By all means question, criticise and make up your own mind. Anyone working in the field of computer game and multimedia development that involves sound will not only be aware of the technical, computing, and engineering issues of their field (the logical ones) but they will doubtless have opinions and their own tastes and creativity. If there is a lesson to be learned from this chapter, it is that we hope you will consider the bigger picture, the wider implications, the external factors and the notion of a user-centred design process. We feel that the three areas of emotion, content and context epitomise these views and will be the crucial issues in future technological and entertainment areas. Think big. In our opinion, the 'blue sky' and 'off the wall' ideas are some of the most fun and interesting things you can do when it comes to being creative with technology. Have fun with your work and work with fun stuff!

REFERENCES

Aucouturier, J. J., & Pachet, F. (2002). Scaling up music playlist generation. In *Proceedings of the IEEE International Conference on Multimedia Expo*.

Bannister, D., & Mair, J. M. M. (1968). *The evaluation of personal constructs*. London: Academic Press.

Battle of the bands. (2008). Planet Moon Studios.

BioShock. (2007). Irrational Games.

Bordwell, D., & Thompson, K. (2004). *Film art: An introduction* (7th ed.). New York: McGraw-Hill.

Clarkson, B., Mase, K., & Pentland, A. (2000). Recognizing user context via wearable sensors. In *Proceedings of the Fourth International Symposium of Wearable Computers*.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, *16*(7/8), 555–575. doi:10.1080/08839510290030390

Cunningham, S., Bergen, H., & Grout, V. (2006). A note on content-based collaborative filtering of music. In *Proceedings of IADIS - International Conference on WWW/Internet*.

Cunningham, S., Caulder, S., & Grout, V. (2008). *Saturday night or fever? Context aware music playlists.* In *Proceedings of the Audio Mostly Conference on Interaction with Sound*.

Cunningham, S., Grout, V., & Hebblewhite, R. (2006). Computer game audio: The unappreciated scholar of the Half-Life generation. In *Proceedings of the Audio Mostly Conference on Sound in Games*.

Dance dance revolution. (1998). Konami.

Dave mirra freestyle BMX. (2000). Z-Axis.

Davies, G., Cunningham, S., & Grout, V. (2007). *Visual stimulus for aural pleasure*. In *Proceedings of the Audio Mostly Conference on Interaction with Sound*.

Davis, H., & Silverman, R. (1978). *Hearing and deafness* (4th ed.). Location: Thomson Learning.

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2003). *Human computer interaction* (3rd ed.). Essex, England: Prentice Hall.

Ekman, I. (2008). Psychologically motivated techniques for emotional sound in computer games. In *Proceedings of the Audio Mostly Conference on Interaction with Sound*.

F.E.A.R. First encounter assault recon. (2005). Monolith Productions.

FIFA. (1993-). EA Sports

Foote, J. (1999). Visualizing music and audio using self-similarity. *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, 77-80.

Freeman, D. (2004). Creating emotion in games: The craft and art of emotioneering™. *Computers in Entertainment*, *2*(3), 15. doi:10.1145/1027154.1027179

Gasser, M., Pampalk, E., & Tomitsch, M. (2007). A content-based user-feedback driven playlist generator and its evaluation in a real-world scenario. In *Proceedings of the Audio Mostly Conference on Interaction with Sound*.

Grand theft auto. (1993-). Rockstar Games.

Grimshaw, M., Lindley, C. A., & Nacke, L. (2008). Sound and immersion in the first-person shooter: Mixed measurement of the player's sonic experience. In *Proceedings of the Audio Mostly Conference on Interaction with Sound*.

Guitar hero. (2005-). [Computer software]. Harmonix Music Systems (2005- 2007)/ Neversoft (2007-).

Hitchcock, A. (Director) (1960). *Psycho*. Hollywood, CA: Paramount.

Jansz, J. (2006). The emotional appeal of violent video games. *Communication Theory*, *15*(3), 219–241. doi:10.1111/j.1468-2885.2005.tb00334.x

Johnstone, T. (1996). Emotional speech elicited using computer games. In *Proceedings of Fourth International Conference on Spoken Language (ICSLP96).*

Jørgensen, K. (2011). Time for new terminology? Diegetic and non-diegetic sounds in computer games revisited . In Grimshaw, M. (Ed.), *Game Sound Technology and Player Interaction: Concepts and Developments*. Hershey, PA: IGI Global.

Károlyi, O. (1999). *Introducing music*. Location: Penguin.

Kelly, G. A. (1955). *The psychology of personal constructs*. New York: Norton.

Kromand, D. (2008). Sound and the diegesis in survival-horror games. In *Proceedings of the Audio Mostly Conference on Interaction with Sound*.

Livingstone, S. R., & Brown, A. R. (2005). Dynamic response: Real-time adaptation for music emotion. In *Proceedings of the Second Australasian Conference on Interactive Entertainment*.

Logan, B. *(2002)*. Content-based playlist generation: Exploratory experiments, *In* ISMIR2002, 3rd International Conference on Musical Information (ISMIR).

Maurizio, V., & Samuele, S. (2007). Low-cost accelerometers for physics experiments. *European Journal of Physics*, *28*, 781–787. doi:10.1088/0143-0807/28/5/001

Moore, B. C. J. (Ed.). (1995). *Hearing: Handbook of perception and cognition* (2nd ed.). New York: Academic Press.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). New York: Academic Press.

Nacke, L., & Grimshaw, M. (2011). Player-game interaction through affective sound . In Grimshaw, M. (Ed.), *Game Sound Technology and Player Interaction: Concepts and Developments*. Hershey, PA: IGI Global.

Parker, J. R. & Heerema, J. (2008). Audio interaction in computer mediated games. *International Journal of Computer Games Technology.*

Platt, J. C., Burges, C. J. C., Swenson, S., Weare, C., & Zheng, A. (2002). Learning a Gaussian process prior for automatically generating music playlists. *Advances in Neural Information Processing Systems*, *14*, 1425–1432.

Pong. (1972). Atari Inc.

Ravaja, N., Saari, T., Laarni, J., Kallinen, K., Salminen, M., Holopainen, J., & Järvinen, A. (2005). The psychophysiology of video gaming: Phasic emotional responses to game events. In *Proceedings of DiGRA 2005 Conference: Changing Views - Worlds in Play*.

Reynolds, G., Barry, D., Burke, T., & Coyle, E. (2007). Towards a personal automatic music playlist generation algorithm: The need for contextual information. In *Proceedings of the Audio Mostly Conference on Interaction with Sound*.

Rock band. (2005-2007). Harmonix Music Systems.

Schmidt, A., & Winterhalter, C. (2004). User context aware delivery of e-learning material: Approach and architecture. *Journal of Universal Computer Science*, *10*(1), 38–46.

Silent hill 2. (2001). Konami.

SingStar. (2004). London Studio.

Space invaders. (1978). Taito Corporation.

Stigwood, R., & Badham, J. (Producers). (1977). *Saturday night fever* [Motion picture]. Hollywood, CA: Paramount.

Sykes, J., & Brown, S. (2003). Affective gaming: Measuring emotion through the gamepad. In *Proceedings of Conference on Human Factors in Computing Systems (CHI '03)*.

Tamminen, S., Oulasvirta, A., Toiskallio, K., & Kankainen, A. (2004). Understanding mobile contexts. *Personal and Ubiquitous Computing*, *8*(2), 135–143. doi:10.1007/s00779-004-0263-1

Ultimate band. (2008). Fall Line Studios.

Wii Music. (2008). Nintendo.

Yost, W. A. (2007). *Fundamentals of hearing: An introduction* (5th ed.). New York: Academic Press.

Zhang, T., & Jay Kuo, C. C. (2001). Audio content analysis for online audiovisual data segmentation and classification. *IEEE Transactions on Speech and Audio Processing*, *9*(4), 441–457. doi:10.1109/89.917689

ADDITIONAL READING

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Brewster, S. A. (2008). Nonspeech auditory output . In Sears, A., & Jacko, J. (Eds.), *The human computer interaction handbook* (2nd ed., pp. 247–264). Philadelphia: Lawrence Erlbaum Associates.

Cunningham, S., & Grout, V. (2009). Audio compression exploiting repetition (ACER): Challenges and solutions. In *Proceedings of the Third International Conference of Internet Technologies and Applications (ITA 09).*

Ekman, I., & Lankoski, P. (2009). Hair-raising entertainment: Emotions, sound, and structure in Silent Hill 2 and Fatal Frame . In Perron, B. (Ed.), *Gaming after dark. Welcome to the world of horror video games* (pp. 181–199). Jefferson, NC: McFarland & Company, Inc.

Freeman, D. (2003). *Creating emotion in games*. Indianapolis, IN: New Riders.

Grimshaw, M. (2008). *The acoustic ecology of the first-person shooter: The player, sound and immersion in the first-person shooter computer game*. Saarbrücken, Germany: VDM Verlag Dr. Mueller.

Grimshaw, M. (2009). The audio Uncanny Valley: Sound, fear and the horror game. In *Proceedings of the Audio Mostly Conference on Interaction with Sound.*

Loy, G. (2006). *Musimathics: The mathematical foundations of music* (*Vol. 1*). Cambridge, MA: MIT Press.

Loy, G. (2007). *Musimathics: The mathematical foundations of music* (*Vol. 2*). Cambridge, MA: MIT Press.

Papworth, N., Liljedahl, M., & Lindberg, S. (2007). Beowulf: A game experience built on sound effects. In *Proceedings of the 13th International Conference on Auditory Display (ICAD)*.

Röber, N., & Masuch, M. (2005). *Leaving the screen: New perspectives in audio-only gaming*. In *Proceedings of 11th International Conference on Auditory Display (ICAD).*

KEY TERMS AND DEFINITIONS

Content: The definable qualities and characteristics for any given piece of information.

Context: The scenario and environment in which a user or application is placed in.

Emotional Interaction: A digital system capable of inducing emotional reactions in a user and being able to dynamically respond to human emotional states.

Emotional Reaction: A human affective response or feeling in response to one or more stimuli.

Emotional State: The dominant, overriding emotional sensation of a human at a given moment.

Playlist Generation: The production of a sequence of songs, often to be listened to on a portable music player.