Kostas Karpouzis
Georgios N. Yannakakis   *Editors*

Emotion in Games

Theory and Praxis

Springer

Socio-Affective Computing

Volume 4

Series Editor
Amir Hussain, University of Stirling, Stirling, UK

Co-Editor
Erik Cambria, Nanyang Technological University, Singapore

This exciting Book Series aims to publish state-of-the-art research on socially intelligent, affective and multimodal human-machine interaction and systems. It will emphasize the role of affect in social interactions and the humanistic side of affective computing by promoting publications at the cross-roads between engineering and human sciences (including biological, social and cultural aspects of human life). Three broad domains of social and affective computing will be covered by the book series: (1) social computing, (2) affective computing, and (3) interplay of the first two domains (for example, augmenting social interaction through affective computing). Examples of the first domain will include but not limited to: all types of social interactions that contribute to the meaning, interest and richness of our daily life, for example, information produced by a group of people used to provide or enhance the functioning of a system. Examples of the second domain will include, but not limited to: computational and psychological models of emotions, bodily manifestations of affect (facial expressions, posture, behavior, physiology), and affective interfaces and applications (dialogue systems, games, learning etc.). This series will publish works of the highest quality that advance the understanding and practical application of social and affective computing techniques. Research monographs, introductory and advanced level textbooks, volume editions and proceedings will be considered.

More information about this series at http://www.springer.com/series/13199

Kostas Karpouzis • Georgios N. Yannakakis
Editors

Emotion in Games

Theory and Praxis

Editors
Kostas Karpouzis
Institute of Communication
   and Computer Systems
National Technical University of Athens
Zographou, Greece

Georgios N. Yannakakis
Institute of Digital Games
University of Malta
Msida, Malta

Foreword

One day, years ago, I unexpectedly found myself several hundred feet below the ocean. My family had been taken from me, and I was hoping that I might be able to find answers deep within a cave riddled with jellyfish and a looming giant octopus. I met an array of sea creatures in otherworldly landscapes. I felt burning curiosity, an inescapable sense of awe and mystery, and indescribable respect for the ocean and the life within. It turns out that my family had been abducted by aliens (we've all been there) who I eventually defeated and returned everyone to safety. Years after I finished playing "Ecco the Dolphin," that experience stayed with me and left a lasting impact. To this day, I have an insatiable love for the ocean and almost became a marine biologist. I was incredibly moved and inspired by my experience with this low-resolution, 16-bit journey.

Art is something that elicits an emotional response – be it music, a painting, a ballet performance, or even video games. Countless modern gamers like myself can tell their own personal story about how they were moved to tears or inspired to action based on a deeply personal experience they had while playing a game.

Video games offer something that almost no other medium can – a two-way dialogue with its participant. A game experience can influence the player and, unlike "conventional" art like paintings, film, or music, the player can influence the experience right back. This two-way street of communication has the power and potential to create an emotional experience that can be more personal, evocative, and effective than any other medium. Furthermore, this dialogue between player and game can become even more powerful when amplified with emerging technologies such as virtual reality, biofeedback, neurofeedback, haptic responses, and whatever new augmentation may be just around the bend.

While this intimate aspect of the "gaming experience" has generally been implicitly understood and casually recognized, the emotional impact and capacity of gaming have only begun to be fully explored and understood.

The chapters within this book examine many of the different and diverse facets of emotion in gaming – what does emotion in gaming even mean, how can it be "accomplished," and why does it all even matter? These are the questions that need

to be answered and the topics that need to be understood before we can boldly go down a path that will open up new ways that games and art, can touch and inspire us to greatness (or, minimally, to become dolphin nerds).

Flying Mollusk, Glendale, CA, USA   Erin Reynolds

Preface

Play constitutes one of our first deliberate activities in life, providing the first opportunity for interaction with objects, devices and other children or grown-ups, months before walking, talking or acquiring advanced tactile abilities. Later on in life, play becomes the basis of forming small or larger groups and identifying rules which need to be respected in order to function within those groups: players of a school basketball team have different roles and competencies, and they all have to work towards the common good of the team, while following the rules of the game, in order to succeed. In this framework, play, especially in social situations, becomes a powerful inducer of emotions, based on its outcome and the dynamics between team members and opponents. In the case of digital games, devices have now the ability to sense player emotion, through cameras, microphones, physiological sensors and player behaviour within the game world, and utilise that information to adapt gameplay accordingly or generate content predicted to improve the player experience and make the game more engaging or interesting.

This book attempts to encompass the concepts that make up the big picture of emotion for and from digital gaming, starting from psychological concepts and approaches, such as modelling of emotions and player experience, to sensing and adapting to emotion, and to discuss applications of emotion-aware gaming in selected concepts and robotics. As work on emotion in games is highly multidisciplinary by nature and important to several areas of research and development, we carefully selected and invited scholars with pioneering work and recognised contributions in game studies, entertainment psychology, affective computing, game design, procedural content generation, interactive narrative, robotics, intelligent agents, natural interaction and interaction design. The result is a **holistic** perspective over the area of emotion in games as viewed by the variant research angles offered by the different research fields. Based on the received chapter contributions, this book is divided in three main parts: *Theory*, *Emotion Modelling and Affect-Driven Adaptation* and *Applications*.

Bateman opens the first part of the book (**Theory**) by examining the question of why we like to play in "The Aesthetic Motives of Play". Kivikangas, in his chapter

titled "Affect Channel Model of Evaluation in the Context of Digital Games", combines different emotion theories, attempting to explain different aspects of game experience. Finally, Calleja et al. discuss "Affective Involvement in Digital Games" through different dimensions of both involvement and games, aiming to introduce affect in their model and examine how different components of gameplay relate to each other and affect player involvement.

The second part of the book (**Emotion Modelling and Affect-Driven Adaptation**) concentrates on computational concepts related to sensing, emotion modelling and generating emotions and content for digital games. Kotsia et al. in "Multimodal Sensing in Affective Gaming" discuss how different sensing modalities (facial expressions, body movement, physiological measurements or even wearables) can provide meaningful insights about the player's emotion to the game, based on player expressivity, while Schuller discusses how speech contributes to player-player and player-game interaction in "Emotion Modelling via Speech Content and Prosody – in Computer Games and Elsewhere". In the framework of brain-computer interfaces and physiology, Fiałek and Liarokapis investigate the use of BCI devices in gaming and virtual reality in "Comparing Two Commercially Available Brain-Computer Interfaces for Serious Games and Virtual Environments" and Yannakakis et al. argue about the importance of physiology for the investigation of player affect in "Psychophysiology in Games". From the synthesis and generation point of view, Ravenet et al. in "Emotion and Attitude Modelling for Non-player Characters" describe how modelling and generation of emotions and attitudes in embodied conversational agents (ECA) can enhance the realism and interactivity of non-player characters. In a similar framework, Togelius and Yannakakis discuss game level design based on affect and other aspects of player experience in "Emotion-Driven Level Generation"; O'Neill and Riedl present Dramatis, a model of suspense used to generate stories that elicit emotions in "Emotion-Driven Narrative Generation"; Burelli crosses the border of traditional cinematography and digital gaming in "Game Cinematography: From Camera Control to Player Emotions"; and Garner describes how sound in games can be utilised to both infer and evoke emotion in his "Sound and Emotion in Video Games" chapter. This part of the book is concluded with the chapter of Broekens et al. "Emotional Appraisal Engines for Games"; the chapter discusses the rationale for specialised emotional appraisal engines for games which provide basic emotion modelling capabilities for the generation of emotions for non-player characters.

The third part of this book (**Applications**) discusses a number of applications that utilise game- and affect-related concepts. Bianchi-Berthouze and Isbister examine body movement as a means for expressing emotion in the context of digital games ("Emotion and Body-Based Games: Overview and Opportunities"). Holmgård and Karstoft in their "Games for Treating and Diagnosing Post-Traumatic Stress Disorder" chapter discuss the capacity of games for mental health and focus on the impact of emotion on games for post-traumatic stress disorder. Khaled et al. present how emotions were used to model conflict, cooperation and competition in a serious game for children, in "Understanding and Designing for Conflict Learning Through Games". Finally, robotic applications of emotion and affect are discussed

by R. Aylett in "Games Robots Play: Once More, with Feeling" from the point of view of robotic companions and digitised games, while Cheok et al. present the fascinating concept of "Lovotics: Love and Sex with Robots".

The editors of this book would like to sincerely thank all contributors for their timely and proactive cooperation and Marleen Moore and all the staff at Springer for their confidence in this project and their support.

Zographou, Greece   Kostas Karpouzis
Msida, Malta   Georgios N. Yannakakis

Contents

Contributors

Ruth Aylett MACS, Heriot-Watt University, Edinburgh, Scotland, UK

Chris Bateman University of Bolton, Bolton, UK

International Hobo, Manchester, UK

Nadia Bianchi-Berthouze UCLIC University College London, London, UK

Rafael Bidarra Intelligent Systems Department, Delft University of Technology, Delft, The Netherlands

Joost Broekens Intelligent Systems Department, Delft University of Technology, Delft, The Netherlands

Paolo Burelli Aalborg University, København, Denmark

Gordon Calleja Institute of Digital Games, University of Malta, Msida, Malta

Adrian David Cheok City University London, London, UK

Imagineering Institute, Nusajaya, Malaysia

Mathieu Chollet LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, Paris, France

Szymon Fiałek HCI Lab, Faculty of Informatics, Masaryk University, Brno, Czech Republic

Maurizio Garbarino Empatica, Milan, Italy

Tom A. Garner School of Creative and Cultural Industries, University of Portsmouth, Portsmouth, UK

George Goudelis Image, Video and Multimedia Systems Lab, National Technical University of Athens, Athens, Greece

Laura Herrewijn Department of Communication Sciences, Research Group CEPEC, Ghent University, Ghent, Belgium

Christoffer Holmgård Department of Computer Science and Engineering, New York University, New York, NY, USA

Eva Hudlicka College of Computer Science, University of Massachusetts-Amherst & Psychometrix Associates, Amherst, MA, USA

Katherine Isbister Computational Media, University of California, Santa Cruz, CA, USA

Richard Joiner University of Bath, Bath, UK

Kostas Karpouzis Institute of Communication and Computer Systems, National Technical University of Athens, Zographou, Greece

Karen-Inge Karstoft Forsvarets Veterancenter, Ringsted, Denmark

Kasun Karunanayaka Imagineering Institute, Nusajaya, Malaysia

Rilla Khaled Concordia University, Montreal, QC, Canada

J. Matias Kivikangas Department of Information and Service Economy, Aalto University, Helsinki, Finland

Irene Kotsia Department of Computer Science, Middlesex University, London, UK

David Levy Imagineering Institute, Nusajaya, Malaysia

Retro Computers Ltd, London, UK

Fotis Liarokapis HCI Lab, Faculty of Informatics, Masaryk University, Brno, Czech Republic

Hector P. Martinez Institute of Digital Games, University of Malta, Msida, Malta

Brian O'Neill Department of Computer Science and Information Technology, Western New England University, Springfield, MA, USA

Ioannis Patras School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK

Florian Pecune LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, Paris, France

Catherine Pelachaud LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, Paris, France

Karolien Poels Department of Communication Studies, Research Group MIOS, University of Antwerp, Antwerp, Belgium

Brian Ravenet LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, Paris, France

Mark Riedl School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, USA

Björn Schuller Imperial College London, London, UK

Julian Togelius Department of Computer Science and Engineering, New York University, New York, NY, USA

Asimina Vasalou Institute of Education, London, UK

Georgios N. Yannakakis Institute of Digital Games, University of Malta, Msida, Malta

Stefanos Zafeiriou Department of Computing, Imperial College London, London, UK

Part I
Theory

Chapter 1
The Aesthetic Motives of Play

Chris Bateman

Abstract Why do people enjoy playing games? The answer, in its most general form, is that there are aesthetic pleasures offered by games and other play experiences that meet powerful and profound human and animal needs. As such, we can identify specific aesthetic motives of play, and one of the clearest ways of characterizing these motives is in terms of the emotional experiences associated with them.

Introduction

If players all wanted the same things, the job of the game designer would be substantially easier, but markedly less rewarding. While there is much to be said for specializing in a particular design practice, most professional game designers do not wish to be tied down to a specific genre of game and as such are interested in the problems of game design as a wider set of problems to be explored. Yet there is no one solution to this problem, and never can be, in part because the aesthetic values that different players hold are radically different [1]. Rather, it is necessary for game designers (and indeed, anyone who studies play) to accept that we are always dealing with a diversity of desires and preferences, whether we characterize this in terms of player types [2, 3], varieties of fun [4], patterns of play [5, 6], play styles [3, 7], or aesthetic values for play [1].

This point is distinct from Brian Sutton-Smith's observation that there is an ambiguity to how we understand play [8], which is a question of interpretation. What is at task here is the question of the *motives for play*, which is partly an empirical question [9] and partly a theoretical one [1]. The intertwining of theory and observation presents a general problem for research in psychology and all the social sciences, since how we investigate a psychological topic, such as motivations, depends to a great degree upon the model we have chosen for doing so [10]. Rather than taking this as an insurmountable problem, we can recognize that the uncertainties of this

C. Bateman (✉)
University of Bolton, Bolton, UK

International Hobo, Manchester, UK

topic are an opportunity for deeper understanding precisely because we can come at the issues from numerous directions. As Whitehead observes, every clash is a sign of "wider truths and finer perspectives", the reconciliation of which can increase our understanding of the complexities of any subject ([11], p. 185).

The purpose of this chapter is thus to ask: *what are the motives of play*? This is also a way of asking: *why do we play*? – provided this latter question is recognized as having more than one response. This is far from the first study to connect motive with play (e.g. [44, 45]), but by synthesizing the work of several research efforts – including empirical, theoretical, and philosophical methodologies – I hope to provide both a summary of what has already been observed, gesture towards new research opportunities, and provide ideas to spur game designers to think in new ways. The answers I will offer here are organized around emotions, but it is not my intention to endorse any particular psychological model of motivation, such as Weiner's attribution model [12]. Such further connections could be added, should there be a need, but my purpose here is more general being focused on philosophical reflections upon the diversity of play, intended to illuminate an otherwise chaotic landscape.

While the source of the multiplicity discussed here are my observations and empirical investigations into play (e.g. [3, 7, 9]), philosophy, in its role as a creator of theories [13], is perfectly suited to analyzing such data. The impression that the sciences are the only fields that should be theorizing in this way conflates the philosophical elements of the scientific practices with their empirical practices: we would be wise to keep in mind that all the sciences were once part of the domain of 'natural philosophy', and have in no way lost their philosophical elements in the years since the 'divorce' between philosophy and empirical science [14].

It is to emphasize the role of philosophy in what follows that I have entitled this chapter 'The Aesthetic Motives of Play', drawing attention to the relationship between games and philosophy of art and aesthetics, as I have done elsewhere in my work [15]. This will be important in what follows, since I will not assert any given definition of 'game' preferring instead to disavow 'game' in order to better understand how games (whatever they are!) are played [16]. This is a quintessentially philosophical reluctance! But it radically improves our chances of getting hold of the widest possible perspective relevant to the given enquiry, and this is precisely the merit of philosophical practices when they work in alliance with empirical research: we are less likely to become trapped inside a single box when we are actively working to construct new boxes, however provisional the method of their construction.

Cautions About Asking 'Why'

To ask 'why do we play?' is a noble question with an honorable tradition behind it. Sutton-Smith's work is demonstrative of the problems of exploring this issue, since different 'rhetorics' will lead to different conclusions [8]. We ought, however, be especially careful that in asking 'why' we do not bring in complications that

will unduly retard our progress in understanding intricate issues. Chief among those we are likely to encounter at this time is the danger of deploying a naïve teleological understanding of evolutionary theories such that the complexity of genuine behavior is trammeled into a compelling yet utterly untestable story about selective advantages or fitness. We can and should question such postulation, especially since in its most common form it is oddly reminiscent of nineteenth century teleological explanations of biology in terms of theology [17].

The issue here is that explaining the properties of (say) animal wings by reference to the equations of aerodynamics represents a fairly strong causal scenario for discussing Darwin's 'descent with modification' [18]. Discussing motivations, however, does not: no matter what situation we are dealing with, the neurobiological functions that we can identify come into play in radically diverse situations, none of which are amenable to expression in terms of simple mathematical equations. Even if we suspect that play is so widespread in animals because it offers survival advantages (which is a plausible hypothesis, asserted by many researchers e.g. [19]), we cannot then reason from this to the biology of any of the neural and biochemical systems associated with specific emotions and behaviors. We not only know too little of the developmental history, but every single emotion has influence on numerous situations relevant to survival and reproduction in ways that makes extracting compact and testable hypothetical scenarios effectively impossible.

The answer to 'why do we play?', therefore, must not be an open door to teleological speculation. Rather, it must invite us to examine patterns in the kinds of play experience different players choose to engage in. It demands that we root our conclusions in observations of the here-and-now, rather than an imagined history of our species at geological time scales. To do so is not to deny our evolutionary history, but simply to recognize the insurmountable limitations in talking about specific aspects of that history, and hence the superior epistemic value of claims that remain focused upon what can be adequately investigated today. Exercising this caution allows us to adequately consider play in non-human animals (which we can also observe today) without drawing erroneous conclusions as to what these parallels embody, thus helping to illuminate the wider significance of the aesthetic motivations of play for the phenomena of life.

Uncertainty as the Foundation of Play

Since I am disavowing 'game', I cannot use a definition of game as a foundation for the investigation that follows. Instead, I shall bound this enquiry upon *uncertainty*, which was first explicitly recognized as having a role in play by Johan Huizinga [23] and Roger Callois [5]. In anthropology, Thomas Malaby has disconnected play from games by suggesting that 'play' is best understood as "a dispositional stance towards the indeterminate" ([20], p. 209). This perspective also aligns with Suits concept of a lusory attitude [21], which is also a state of mind adopted for the purpose of play. Suits, however, constrains his perspective on games to cases involving "bringing about a specific state of affairs" ([22], p. 148) i.e. to goal-

oriented activities. Malaby's perspective uncouples 'game' (as a kind of activity) from play (as a mental state) – freeing the study of games and play from many restrictive assumptions about what a game must be.

This 'liberation of games', I have argued, is a necessary corrective to the trend in game studies to restrict interest solely to certain narrowly construed cases [24]. It allows us to examine a variety of related cases where play occurs but which might not usually be considered games, and in so doing brings attention to the ways that other human practices – such as narrative – can be understood as being actively played [25]. Doing so also helps draw attention to the way that specific aesthetic values in relation to play become asserted as attempts to define boundaries for terms like 'game', limiting attention to some subset of the motives I will outline in what follows [1].

In order to structure the discussion of the motives of play, I will treat as distinct those motives that emerge from the *representational* aspects of play from those that emerge from the *functional*, i.e. Jesper Juul's fiction vs. rules [26]. It is worth noting, however, that the functional elements (rules) could also be subsumed in the fiction (as Walton's prescriptions to imagine [27]) just as the fiction could be subsumed in the rules, depending on what is taken as ontologically primitive [15]. The distinction helps separate out goal-oriented motives from alternatives [28], but should not be mistaken for anything deeper than a general organizing principle. To begin with, however, it is worth identifying three general aesthetic motives that can be expressed in play.

General Motives

The three motivations for play discussed in this section do not clearly resolve into uniquely functional or representational roles. Rather, they possess elements of both, or manifest in one fashion for some players and differently for others. All three have in common that the underlying biology behind the behaviors in question supports extremely general aspects of lived existence; as a result, these are perhaps the most general motivations for engaging in play – indeed, they are extremely general motivations for pursuing any activity.

The Social Motive

If we take Huizinga as the beginning of the study of play and games, then from the very beginning it was recognized that play was at heart a social phenomena [23]. Indeed, Huizinga's concerns that the play element in culture was declining are based around the recognition that the increasing importance assigned to the moral and aesthetic value of efficiency was driving out play as the foundation of culture. Similarly, the moment there were player models exploring the diversity of play there was a recognition of the social motive, as indicated by Richard Bartle's Socializer

type [2], and the BrainHex play style descended from it [47]. More recently, a social aesthetic for play has been identified by considering definitions of 'game' [1].

The scope and range of the social motive for play is too vast to be adequately summarized, but it is important to recognize that it manifests in both functional and representational forms. As an example of the former, the delight in *schadenfreude* (pleasure in the misfortune of others) [30] is something inherent to multiplayer games of all kinds [29] but it emerges from the jostling for victory (i.e. the winning state, a functional element of play). The social motive can also be found in a purely representational form, as can be seen in case studies of tabletop role-players, especially in freeform games [51], and in many of the artgames of Tale of Tales such as *The Endless Forest* [100] or *Bientôt L'été* [101] that create unique encounters between people in an artistically-motivated fictional world. Furthermore, even though some contemporary game forms seem to avoid the social motive – single player videogames, for instance – studies of players reveal indirect social elements to play, such as when groups of friends opt to play the same single player game in parallel [3].

Needless to say, this social aesthetic motive is not uniquely human, and is easily observed in familiar animals such as dogs and cats, as well as in birds, lizards, and perhaps even fish [52]. Every species expresses social motivations within play in different ways, but there is often substantial cross-over: a wolf and a coyote can play together, because they share common body language for initiating play. Similarly, humans can play with dogs and cats without any difficulty, and indeed cats and dogs (and cockatoos, and rats, and . . .) can play together under appropriate circumstances [53].

Oxytocin in mammals, and equivalent hormones in other species (such as meso-tocin in birds [59]), provide the neurobiological foundation for social motivations [54, 55], and oxytocin levels have been shown to be elevated in both humans and animals when they live and play together [56–58]. It is important to remember, however, that behavior is not reducible to its chemical substrate: indeed, the sheer variety of behaviors facilitated by this aspect of biology should make this caution quite clear.

The Thrill-Seeking Motive

Another extremely general aesthetic motive is the pursuit of excitement, the neuro-chemical substrate for which is epinephrine [60], commonly called 'adrenalin'. This aesthetic motive is even older than the social motive, as indicated by recent evidence of 'thrill-seeking' among bees [61]. As with the social motive and oxytocin, the purpose of mentioning this alternative level of observation isn't to replace behavior with low-level biology: on the contrary, we cannot substitute actual behavior for biological substrates since we do not have a complete picture of any phenomena until we have studied it from every possible perspective.

Caillois was the first to identify this motive in his *ilinx* (vertigo) pattern [5], which has been expressly linked to excitement and epinephrine [6]. Nicole Lazzaro notes the relationship between excitement and relief in players of videogames [4], and the same pattern reappears in the Daredevil play style in BrainHex [47]. Because almost all kinds of play elicit excitement in one form or another, it is important to be clear that the thrill-seeking motive is intended to mark an inducement for play distinct from the pursuit of victory or other functional motives: when a speed-runner challenges themselves to complete a videogame as fast as possible without saving (thus upping the risks of failure), their motive can be understood as thrill-seeking in the sense intended here (see [105, 106]). Similarly, players who engage in snowboarding simply for the vertigo of descending the mountain can be understood as motivated by the aesthetic values sketched here [62].

The Curiosity Motive

The third and final general aesthetic motive for play shares with the other two the utter generality of the behavioral pattern and the ancient origins of its neurobiological substrate. Curiosity (known in the psychological literature as 'interest') has been linked to the neurotransmitter endomorphin [63], as well as to the most general motivation chemical, dopamine [64] (although discussion of the biology of motivation in regard to dopamine is beyond the scope of this discussion, precisely because it is far too general). Irving Biederman and Edward Vessel draw attention to our fascination with 'richly interpretable' patterns – which includes, in practice, everything from stunning mountaintop vistas to mathematical problems.

Although one of the key figures in the psychology of emotions, Paul Ekman, does not interpret curiosity as an emotion [30], it is treated by some researchers (such as Lazzaro [4, 29]) as emotion-like, and there is a growing trend towards considering it as an emotion [73]. However, curiosity can also be connected to another of Ekman's emotions: *wonder*. This is a powerful full-body emotion, as strong as the emotional experience acquired through victory (i.e. *fiero* – see below) [29]. Since no-one can directly anticipate that they will encounter an experience of wonder in any given context, it is plausible to consider the curiosity motive as entailing the pursuit of wonder as part of its broadest definition.

The appearance of wonder in games is often coupled with fear to produce awe, as for example in the case of the titular creatures in *Shadow of the Colossus* [102], that are expressly designed to create a wondrous sense of scale to enhance the player's terror in facing them. However, wonder can also be elicited when the design allows for it, as with the set pieces in *Endless Ocean* [103] whereby whales and other large aquatic animals are made to appear to the player at certain places in the game world. Experiences such as this blend wonder with a sense of the beauty of nature, much discussed by Kant [104] and other philosophers. Indeed, this connection may go deeper: it has been suggested that the joy to be found in beauty – one of the

most quintessential aesthetic pleasures – could be founded upon curiosity [70, 71], making this motive exceptionally important for a wide variety of artworks, including videogames.

It is likely that the curiosity motive is present in a great deal of play, but that other aesthetic motives 'drown it out', making it difficult to detect (especially in competitive play). Nonetheless, since the 1980s there has been a clear recognition of the importance of curiosity as an aesthetic motive, especially in the work of Thomas Malone regarding what makes something fun to learn [65], and what makes an enjoyable user interface [66]. Curiosity has been linked to play styles in the Wanderer play style [3], Lazzaro's Easy Fun [4], and the Seeker play style in BrainHex [47]. Furthermore, the curiosity motive is far from unique to humans, having been observed in a wide range of species including rats [67], ravens [68], and even octopi [69]. Life is curious, and this curiosity significantly affects how animals play.

Functional Motives

The previous motives are extremely general, and as such while they do constitute aesthetic motivations for play, they are also motivations in a more general sense. However, whenever someone is motivated to play by something arising from the structural character of play (the rules, or practices that can be efficiently characterized using rules), we can identify patterns in their motives expressible in terms of relations between emotions. Such *functional motives* have in common the relevance of defined states in the play space being experienced – states such as 'winning', 'losing' and '100 % complete'. In such cases, it is usually not ambiguous that we are dealing with a game i.e. few definitions of 'game' would exclude such cases from consideration. As such, these are the most explicitly game-like motives, although this should not be used to construe that these alone adequately characterize games, for reasons that go beyond the scope of this chapter (see [24] for detailed discussion of this point).

The Victory Motive

Everybody likes to win, but not everyone is equally motivated to achieve success in the games they play. Certain players, however, are specifically attracted to challenge – to the extent that they are far more likely to complain about games being 'too easy' than 'too hard' [3]. As Lazzaro reports [29], players expressly focused upon earning victory are seeking the emotional reward Ekman characterizes as *fiero* [30], which can also be termed triumph [15], the clear sign of which is hands held aloft or a 'fist pump', perhaps accompanied by a mouth-opened, teeth-bared 'barbaric yawp' [31].

The characteristic of this kind of play is that the final emotional reward is not just victory, per se, but triumph over adversity – hence the problem with games that are 'too easy': if there is no perceived struggle to attain success, the emotional payoff of fiero is not attained. As a result, the victory motive typically also entails experiences of frustration and hence of anger prior to the eventual victory. It is not that anger is *required* to attain fiero, so much as there must be a perception of struggle. In a digital game, a player defeating a boss that has caused them problems in the past will produce the emotional reward of fiero even if a particular attempt at overcoming the boss did not elicit frustration.

This aesthetic motivation has been characterized as the Conqueror play style [3, 9], Hard fun [29], and the victory aesthetic [1], and was first discussed as a unique pattern of play by Caillois as *agon* [5]. The fundamental psychological trait can be characterized as *frustration endurance* [9], which may relate to the hormone testosterone [33–35]. It is the quintessential motivation within all sports, where it is often termed the achievement motive [36]. Although the emotional relationship between fiero and anger at the heart of the victory motive is an aspect of human nature, this motivation can also be observed in predatory mammals and birds. Dogs in a park chasing a ball or playing tug demonstrate the same overall patterns of behavior, as do birds-of-prey when they compete with one another, as indeed was noted by Darwin [32].

The Problem-Solving Motive

Whereas the victory motive seeks triumph in the context of challenge through the endurance of frustration, the same emotional reward is also available through more cerebral pursuits. The solving of puzzles can also trigger fiero, or less intense positive affect emotions [41] such as satisfaction, but in such cases endurance of frustration is not what is typically observed. Rather, incomplete information or conceptual blocks [37] form the typical circumstances by which a puzzle is mounted, requiring those who intend to solve the problem to think the problem through carefully. As a result, this motive can be connected to the putative psychological trait of *confusion endurance* [9]: those who enjoy solving puzzles are content to put up with a state of confusion in a parallel way to the victory motive's endurance of frustration.

The problem-solving motive has been described as the Manager [3] or Mastermind [9] play style, and the problem aesthetic [1]. It is evidenced in the enjoyment of crossword puzzles, sudoku, digital adventure games (i.e. text adventures, point-and-click adventures), and also in the play of tabletop wargames and digital strategy games. In these latter cases, strategic decision making forms a parallel to puzzle-solving in cases where even *formulating* the puzzle to be solved is part of the pleasures of solution. Once again, this motive can be comfortably extended into other animal species: primates [38] and crows [39] are among the species that have

been observed taking pleasure in finding solutions to puzzles. However, humans appear to be the only species that takes a delight in *creating* puzzles for the sake of solving them [40].

The Luck Motive

There is a third aesthetic motive we can associate with the pursuit of triumph, one that is often overlooked in discussions of games and their design: luck. Games of chance were recognized as significant by Caillois, who termed them *alea* [5]. He observed that in a game of skill, the player with the most talent is likely to emerge victorious, but in games of pure chance, *everyone has an equal chance of winning*. This makes games of luck appealing to many players because they do not need to possess any special skills to achieve victory. Indeed, Caillois suggests this helps explain the appeal of lotteries, since those with the misfortune to be born without the talents to succeed in life are offered a 'second chance' of success, irrespective of their circumstances.

Chance is also a significant source of variety in games of many kinds, especially boardgames and videogames with procedural elements [6], but in many of these cases it does not represent a motive for play, as such. For instance, in a procedural sandbox game such as *Minecraft* [99], the fictional worlds are randomly generated but the curiosity motive is a better explanation of the appeal than the luck motive. The luck motive comes to the fore in casino games such as roulette, which are won by pure chance, and similarly in bingo (for which the social motive is also a significant factor of appeal), and of course lotteries of all kinds. In all such cases, it is possible to reach the emotional reward of fiero without typically risking frustration, precisely because failure is not perceived as the player's responsibility [6].

The Acquisition Motive

While the victory and problem-solving aesthetic motives were based around overcoming – of challenges and puzzles respectively – players also gain enjoyment in play from the acquisition of anything at all with a defined role within any given play space. This acquisition motive typically produces a low degree of positive affect such that an observed player may not even show signs of enjoyment on their face. Nonetheless, the act of collecting is an observable motive for play [43], and the act of completion (when the collection in question has a finite scope) can be extremely rewarding. It is unlikely to elicit fiero, but it can produce other positive affect emotions such as satisfaction, which Ekman lists among the happy emotional experiences [42]. The aesthetic motive of acquisition can also produce flow states [46], if the frequency of acquisition is sufficiently frequent.

In some cases, the enjoyment of acquisition produces more than just a general motivation: it can become a driving force behind play, as in Bartle's Achiever type [2] and the BrainHex Achiever play style that descends from it [47]. In such cases, the motive is not merely collecting but *completing*, and the drive to complete can result in players going through highly repetitive play that may be experienced as boring. Whereas other players view excessive boredom as a reason to stop playing a game, those fitting this archetype possess a putative psychological trait termed *boredom endurance* [9], and remain motivated towards the attainment of a perfect state of completion, often measured explicitly by games as a 100 % complete state.

While it is difficult to assess whether non-human animals take the acquisition motive as far as the Achiever play style, the aesthetic motive to acquire can be witnessed in many animal species. Here there might be a danger of bringing in teleological justifications (discussed above) since in many cases the acquisition focuses upon food e.g. larder-hoarding and scatter-hoarding behaviors in squirrels and birds [48]. However, there are clear cases where the aesthetic motive of acquisition cannot be explained as a survival instinct e.g. rats have been observed collecting trinkets [50], and John Endler has argued that male bowerbirds have aesthetic motivations when they collect material to build and decorate their bower [49]. Further ethological research is required here: the interest in providing putative evolutionary explanations may have distorted research on collection behaviors where these represent an aesthetic motive rather than a survival instinct. Additionally, the case discussed by Endler might actually be better considered in terms of other aesthetic motives (such as the curiosity motive mentioned above), particularly if beauty can be considered as an experience associated with curiosity.

Representational Motives

Finally, we come to those aesthetic motives that occur through the representational elements of play. One important distinguishing characteristic of this group of motives is that they are uniquely human (or at least, they are as far as we can ascertain). While other animals do engage with representations in various ways – recall the aforementioned bowerbird – only humans engage in substantial activities of play within fictional worlds of their own devising. That said, it should be acknowledged that animal play can also be understood in terms of fictional worlds [15], and imagination can be extended throughout the history of life provided certain provisos are accepted [72]. Nonetheless, the sophistication of humanity's fictional worlds is matched only by their diversity and ubiquity: we are a species that thrives upon exercising our imagination in many ways. The remaining aesthetic motives are variations upon this peculiar theme.

The Narrative Motive

One of the most striking aspects of human-crafted narratives are their capacity to elicit any and all emotions: we empathize and identify with the protagonists of stories, and vicariously experience their highs and lows whether through the written word, audio plays, or the audio-visual spectacle of television and cinema [74]. This is a topic with a long history in philosophy, going back to the ancient Greek philosophers such as Aristotle [75] and still generating interesting discussion today (e.g. [27]). While this aesthetic motive is undoubtedly rooted upon the same biological substrate as the curiosity motive, it would be misleading to conflate human interest in narrative and narrative-generating systems (including many boardgames and videogames) with curiosity alone. The enjoyment of narrative possesses a complexity that goes beyond simple emotional primitives, a point foreshadowed by Joseph Campbell's recognition of the psychological ubiquity of mythic themes across all cultures [76]. It cannot be taken as a coincidence that Hollywood scriptwriters have been able to reverse engineer screenwriting techniques from Campbell's mythic templates into successful motion pictures [77].

By the *narrative motive*, I seek to mark the motivation for engagement with stories over and above the other aesthetic motives described here. Of course, the curiosity motive plays a key role – but so too does the thrill-seeking motive, and indeed, all the other aesthetic motives. Narratives engage us on many levels, but to take an interest in stories is in itself an aesthetic motivation, it is what makes the reading of novels or the enjoyment of cinema into a lifelong hobby. Disentangling the narrative motive from other motivations in games is so difficult that many researchers simply set narrative aside (e.g. [4]). Nonetheless, it should never be forgotten that the practices of storytelling provide an intrinsic motivation for play *through the properties of narratives and narrative systems themselves*. Indeed, our enjoyment of stories can always be understood as play [25], despite this not being the typical way of interpreting our engagement with novels, movies, and so forth.

The Horror Motive

Why do some people take such pleasure in pursuing apparently unpleasant emotions such as fear and disgust? The matter has generated a substantial literature both in philosophy and in psychology, and indeed in the intersection between the two (e.g. [78]). Noël Carroll has written extensively on the subject, and in the context of the enjoyment of terror denies that it is the fearful experience that provides the appeal, instead reducing this to the curiosity motive [79]. As with so many such examples, curiosity undoubtedly plays a role – but Carroll is perhaps too quick to dismiss enjoyment of fear, since empirical research suggests that awareness that what is happening is fictional creates the conditions for enjoyment [80].

A fascinating and contentious approach to the problem comes from Kendall Walton, who suggests that we need to treat our response to fictional horror as phenomenalogically distinct from fear [81]. While watching a movie, we experience something that seems like terror *and yet we do not run away*. Walton argues that "Fear emasculated by subtracting its distinctive motivational force is not fear at all" [27, p. 201–202]. Walton suggests the term quasi-fear to mark this distinction (and indeed, proposes quasi-emotions in general as categories for dealing with unique emotional responses to fiction). This has generated a substantial debate, nearly summarized by Berys Gaut as consisting of two rival camps – emotional irrealists, such as Walton, Gregory Currie and Jerrold Levinson, and emotional realists, such as Peter Lamarque, Noël Carroll, John Morreall, Derek Matravers and Richard Joyce [82]. Gaut argues that we should not evoke quasi-emotions if the realist account is simpler and adequately explains the relevant phenomena. Yet it is far from clear that this is the case, especially since Gaut (and many others who have commented on this issue) confuse pretending to be afraid with imagining that we are afraid [83].

This issue can be untangled using Tamar Szabó Gendler's concept of *alief* [84, 85] which denotes a more primitive form of belief, one that can be linked to the ancient brain structures known as the limbic system, and specifically the amygdala [15, 72]. If beliefs are understood as having propositional content (e.g. "I believe it is raining"), aliefs are more primitive, being representational but not propositional (we perceive a threat, but at a pre-conscious level). In a horror game, movie, or similar situation we believe we are safe – but we alieve that we are in danger, a situation analogous to the enjoyment of rollercoasters [15]. This is why quasi-fear has the physiological characteristics of fear – we have a neurochemical experience that is empirically indistinguishable from fear.

However, whatever is going on at the level of the biological substrate, the phenomenal character of the situation is indeed radically different: our aliefs and beliefs do not accord in the case of quasi-fear, and this affects our behavior. Thus emotional irrealism wins out, since Gaut's realist position is an incomplete description. Such situations involve what Gendler terms "belief-discordant alief" ([84], p. 641); we alieve we are in danger but we do not believe it. This is Walton's quasi-fear. The enjoyment of this state, and thus its pursuit, forms another aesthetic motive for play, the *horror motive*, which can be associated with the Survivor play style in BrainHex [47]. The related enjoyment of disgust could be explicated by parallel arguments.

The Agency Motive

While our enjoyment of narrative can be characterized as play, and is not as passive as is sometimes suggested [25], there is another expressly aesthetic motive that can be associated with the unique kinds of richly-expressive activities that emerge from our play with toys and games. This *agency motive* places aesthetic value in

the perception of interactivity or the freedom to have a recognizable impact, the paradigm case for which is perhaps children's play with toy building block sets [86]. It is also readily apparent in the aesthetic values asserted in the context of games [1], where it may also be conflated with decision-making (which would be better understood in terms of the problem-solving motive, discussed above). It is also possible that this aesthetic motive could be understood as a specific expression of a more general imaginative aesthetic [1], which may also intersect with the narrative motive, discussed above.

The emotional rewards that can be related to this motive are those of the positive affect emotions such as satisfaction, perhaps even fiero in some cases where an individual strives to express their agency against imposed restrictions. There is little mystery to the idea that exerting control or influence over environment is a pleasant experience, or even an element in well-being, even if understanding agency as a phenomena is anything but simple [94, 95]. Furthermore, despite the Enlightenment tendency to treat agency as uniquely human, there is a growing recognition that the imputation of agency can be extended to other animals [96]. This matter was still widely considered outlandish when discussed by Alasdair MacIntyre as late as 1999 [97], yet it now represents a major research area in philosophy and other fields [98].

Alas, there has been little explicit discussion of agency as an aesthetic motive precisely because the assumption that games can be understood as 'interactive' (as against conventional narrative which is deemed 'passive'). This perspective conflates agency with games, blurring the aesthetic values entailed in valorizing agency. This does not mean agency in games is not discussed – there is a substantial and rich literature (e.g. [87–91]), full of fascinating observations and ideas. But as previously mentioned, the binary division into interactive and passive not only misrepresents the play of stories [25], it also misrepresents the play of games and toys since it is far from the case that all such play activities embody agency, or require an agency motive to be understood or enjoyed. The result is a failure to adequately explore the aesthetic qualities of agency, particularly in the context of enjoyment of games where agency is *absent* (consider *Progress Quest*, for instance [92]), and no-one has yet proposed a play style rooted in agency (perhaps because of the assumption that it is ubiquitous). This is another case where the 'liberation of games' [24] may be a requirement for achieving a greater appreciation for the subtlety and diversity of the aesthetic motives of play.

Conclusion

Caillois was perhaps the first person to look at games and recognize that there were different motives involved in playing them [5]. The aesthetic experiences of sports and other competitive activities (*agon*) are radically different from the surrender to fate (*alea*), the consciousness-disrupting power of vertigo (*ilinx*), and the allure of fiction (*mimicry*). His patterns of play, which were never intended as a taxonomy or

a complete set, can be understood in terms of the emotional experiences invoked, or even the neurobiological substrate [6]. In each case, something more is added into the description, which is never reducible to a single factor; play is always more subtle and complex than such reductive explanations imply. In the discussion above, four of the aesthetic motives of play described descend directly from Caillois' work (the victory motive from agon, the luck motive from alea, the thrill-seeking motive from ilinx, and the narrative motive from mimicry).

The presence of a variety of motives paints an image of play as diverse: this is a point I continually stress within my work, both as a game designer and as one who studies play and games with both an empirical and a theoretical eye. From the general motivating force of the social, thrill-seeking, and curiosity motives; through the game-specific motives of victory, problem-solving, and acquisition, and on into the representative motives of narrative, horror, and agency; the scope of the domain of play is vast – particularly when we take the play of other animals into account as well as that of humans. Precisely because of this immense heterogeneity, many have tried to cleave-off aspects of play and examine them in isolation – as games, as narratives, as art. There is much to be gained from the narrow focus implied by this – provided we do not fool ourselves into taking our boundary fences as something ontologically deeper than tools of convenience.

As Mary Midgley remarked in 1974, games and art must possess a conceptual unity because they "deal with human needs, which certainly do have a structure" ([93], p. 253). But the unity of the concepts of 'game' and 'art' does not mean that our experiences of either can be reduced to simple primitives. If we do so for any reason other than closely examining an intentionally constrained domain of interest, we risk fooling ourselves into thinking that conceptual divisions are something more than conventions. Thus while we can identify a domain of 'games' or of 'art', the aesthetic motives I describe here are equally applicable to either domain – indeed, many of these motives are equally applicable to animals other than humans.

Acknowledging the aesthetic motives of play, whether in the form I outline here or some other arrangement, means both accepting the diversity of human play experience, and also recognizing the relationship between this and animal play. As a game designer, I find this enquiry vertiginous – it hints at the boundless potential of the myriad playful media we are exploring. As a philosopher, I find it awe-inspiring – 4 billion years of life have developed and refined aesthetic motives that we share with many other animals, and that we humans are uniquely positioned to take forward in entirely new and unexpected ways. If I can share any aspect of my wonder in the face of the aesthetic motives of play with others, that would be enough for me to rest content. If in so doing I can add to our appreciation of games, of art, and of the intersection between the two, I will have achieved more than I ever hoped.

References

1. Bateman C (2014) Implicit game aesthetics. Games and Culture. Available online: http://gac.sagepub.com/content/early/2014/12/11/1555412014560607.abstract. Accessed 10 Mar 2015, Article published online before print
2. Bartle R (1996) Hearts, clubs, diamonds, spades: players who suit MUDs. J MUD Res 1(1). [online] http://www.mud.co.uk/richard/hcds.htm. Accessed 5 Mar 2015
3. Bateman C, Boon R (2005) 21st century game design. Charles River Media, Boston
4. Lazzaro N (2003) Why we play: affect and the fun of games. In: Sears A, Jacko JA (eds) The human-computer interaction handbook: fundamentals, evolving technologies, and emerging applications. Lawrence Erlbaum, New York, pp 679–700
5. Caillois R (1958) Man, play and games (trans: Barash M [1962]). Thames & Hudson, London
6. Bateman C (2009) Understand patterns of play. In: Bateman C (ed) Beyond game design: nine steps towards creating better videogames. Charles River, Boston, pp 61–116
7. Bateman C, Lowenhaupt R, Nacke LE (2011) Player typology in theory and practice. Proceedings of DiGRA 2011 Conference: Think Design Play, Utrecht, Netherlands (Sept). Available online: http://www.digra.org/wp-content/uploads/digital-library/11307.50587.pdf
8. Sutton-Smith B (1997) The ambiguity of play. Harvard University Press, Cambridge, MA
9. Bateman C (2014) Empirical game aesthetics. In: Angelides MC, Agius H (eds) IEEE handbook of digital games. Wiley-IEEE Press, New York, pp 411–443
10. Griffin C (2003) The advantages and limitations of qualitative research in psychology and education. Sci Ann Psychol Soc North Greece 2:3–15
11. Whitehead AN (1925/1967) Science and the modern world. Free Press, New York
12. Weiner B (1986) An attributional theory of motivation and emotion. Springer, New York
13. Walton KL (2007) Aesthetics – What? Why? and Wherefore? J Aesthet Art Crit 65(2):147–161, Spring 2007
14. Cahan D (ed) (2003) From natural philosophy to the sciences: writing the history of nineteenth-century science. University of Chicago Press, Chicago
15. Bateman C (2011) Imaginary games. Zero Books, Winchester
16. Bateman C (2015) A disavowal of games. Proceedings of the second Philosophy at Play Conference, University of Glocestershire (forthcoming)
17. Bateman C (2012) The mythology of evolution. Zero Books, Winchester
18. Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favored races in the struggle for life. John Murray, London
19. Smith PK (1982) Does play matter? Functional and evolutionary aspects of animal and human play. Behav Brain Sci 5(1):139–155
20. Malaby TM (2009) Anthropology and play: the contours of playful experience. New Lit Hist 40(1):205–218
21. Suits B (1978) The grasshopper: games, life and utopia. Scottish Academic Press, Edinburgh
22. Suits B (1966) What is a game? Philos Sci 34(2):148–156, June
23. Huizinga J (1949) Homo Ludens : a study of the play-element in our culture (trans: Hull RFC). Routledge & Kegan Paul, London
24. Bateman C (2015) Fiction denial and the liberation of games. *Ab extra* paper at DiGRA 2015
25. Bateman C (2014) What are we playing with? Role-taking, role-play, and story-play with Tolkien's legendarium. Int J Play 3(2):107–118
26. Juul J (2005) Half-real, video games between real rules and fictional worlds. MIT Press, Cambridge, MA
27. Walton KL (1990) Mimesis as make-believe: on the foundations of the representational arts. Harvard University Press, Cambridge, MA

28. Button SB, Mathieu JE, Zajac DM (1996) Goal orientation in organizational research: a conceptual and empirical foundation. Organ Behav Hum Decis Process 67(1):26–48
29. Lazzaro N (2009) Understand emotions. In: Bateman C (ed) Beyond game design: nine steps towards creating better videogames. Charles River Media, Boston
30. Ekman P (2003) Emotions revealed. Times Books Henry Holt and Company, New York
31. Whitman W (1855/2008) "Song of Myself", leaves of grass. Digireads.com, New York, pp 19–63
32. Darwin C (1871) The descent of man, and selection in relation to sex. John Murray, London
33. Andrew RJ, Rogers LJ (1972) Testosterone, search behaviour and persistence. Nature 237(5354):343–346, June
34. Booth A, Shelley G, Mazur A, Tharp G, Kittok R (1989) Testosterone, and winning and losing in human competition. Horm Behaviour 23(4):556–571
35. Elias M (1981) Serum cortisol, testosterone, and testosterone-binding globulin responses to competitive fighting in human males. Aggress Behav 7(3):215–224
36. Cashmore E (2002) Sport psychology: the key concepts. Psychology Press, London
37. Moscovich I (2004) The shoelace problem and other puzzles. Sterling Publishing Company, New York
38. Clark FE, Smith LJ (2013) Effect of a cognitive challenge device containing food and non-food rewards on chimpanzee well-being. Am J Primatol 73:807–816
39. Jelbert SA, Taylor AH, Cheke LG, Clayton NS, Gray RD (2014) Using the Aesop's fable paradigm to investigate causal understanding of water displacement by New Caledonian crows. PLoS ONE 9(3):e92895, available online: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0092895. Accessed 2 Mar 2015
40. Danesi M (2004) The puzzle instinct: the meaning of puzzles in human life. Indiana University Press, Bloomington
41. Plutchik R (1994) The psychology and biology of emotion. HarperCollins College Publishers, New York
42. Ekman P (1992) An argument for basic emotions. Cognit Emot 6(3–4):169–200
43. Bateman C (2009) Include players with different skills. In: Bateman C (ed) Beyond game design: nine steps towards creating better videogames. Charles River Media, Boston, pp 189–212
44. Yee N (2006) The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. Presence Teleop Virt 15(3):309–329
45. Malone TW (1981) Toward a theory of intrinsically motivating instruction. Cogn Sci 5(4):333–369
46. Csikszentmihalyi M (1990) Flow: the psychology of optimal experience. Harper and Row, New York
47. Nacke LE, Bateman C, Mandryk RL (2011) BrainHex: preliminary results from a neurobiological gamer typology survey. Proceedings of 10th International Conference on Entertainment Computing (ICEC 2011). Vancouver, pp 288–293
48. Dally JM, Emery NJ, Clayton NS (2005) Cache protection strategies by western scrub-jays, Aphelocoma californica: implications for social cognition. Anim Behav 70(6):1251–1263
49. Endler JA (2012) Bowerbirds, art and aesthetics: are bowerbirds artists and do they have an aesthetic sense? Commun Integr Biol 5(3):281–283
50. Mitchell J (1959) The bottom of the harbor. Pantheon Books, New York
51. Hughes J (1991) New directions in Australian roleplaying: style and innovation in roleplaying design. Second Roleplaying Forum, Sydney (July). Available online: http://myth-o-logic.org/systemless-roleplaying/1086-2/. Accessed 9 Mar 2015
52. Burghardt GM (2005) The genesis of animal play: testing the limits. MIT Press, Cambridge, MA
53. Kuo ZY (1960) Studies on the basic factors in animal fighting: VII. Inter-species coexistence in mammals. J Genet Psychol 97:211–225

54. van Leengoed E, Kerker E, Swanson H (1987) Inhibition of post-partum maternal behavior in the rat by injecting an oxytocin antagonist into the cerebral ventricles. J Endocrinol 112(2):275–282, February
55. Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U, Fehr E (2008) Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron 58(4):639–650, May
56. Nagasawa M, Kikusui T, Onaka T, Ohta M (2009) Dog's gaze at its owner increases owner's urinary oxytocin during social interaction. Horm Behav 55:434–441
57. Handlin L, Hydbring-Sandberg E, Nilsson A, Ejdebäck M, Jansson A, Uvnäs-Moberg K (2011) Short-term interaction between dogs and their owners: effects on oxytocin, cortisol, insulin and heart rate. An explorative study. Anthrozoös 24:301–315
58. Beetz A, Uvnäs-Moberg K, Julius H, Kotrschal K (2012) Psychosocial and psychophysiological effects of human-animal interactions: the possible role of oxytocin. Front Psychol 3:234
59. Goodson JL, Schrock SE, Klatt JD, Kabelik D, Kingsbury MA (2009) Mesotocin and nonapeptide receptors promote songbird flocking behavior. Science 325:862–866
60. Frijda NH (1986) The emotions. Cambridge University Press, Cambridge
61. Liang ZS, Nguyen T, Mattila HR, Rodriguez-Zas SL, Seeley TD, Robinson GE (2012) Molecular determinants of scouting behavior in honey bees. Science 335(6073):1225–1228
62. Bateman C (2005) Portrait of a type 3 wanderer. Only a Game. Available online: http://onlyagame.typepad.com/only_a_game/2005/09/portrait_of_a_t.html. Accessed 9 Mar 2015
63. Biederman I, Vessel EA (2006) Perceptual pleasure and the brain. Am Sci 94:247–253, May-June
64. Yue X, Vessel EA, Biederman I (2007) The neural basis of scene preferences. NeuroReport 18(6):525–529
65. Malone TW (1980) What makes things fun to learn? Heuristics for designing instructional computer games. Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems. pp 162–169
66. Malone TW (1981) Heuristics for designing enjoyable user interfaces: lessons from computer games. Proceedings of the Conference on Human Factors in Computer Systems (CHI). pp 63–68
67. Berlyne DE (1955) The arousal and satiation of perceptual curiosity in the rat. J Comp Physiol Psychol 48(4):238–246
68. Heinrich B (1999) Mind of the raven: investigations and adventures with wolf-birds. HarperCollins, New York
69. Kuba MJ, Byrne RA, Meisel DV, Mather JA (2006) When do octopuses play? Effects of repeated testing, object type, age, and food deprivation on object play in octopus vulgaris. J Comp Psychol 120(3):184–190, Aug
70. Lehrer J (2011) Why does beauty exist? Wired. Available online: http://aminotes.tumblr.com/post/10845779320/why-does-beauty-exist-jonah-lehrer-beauty-is-a. Accessed 9 Mar 2015
71. Ishizu T, Zeki S (2011) Toward a brain-based theory of beauty. PLoS ONE 6(7):e21852, available online: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021852. Accessed 9 Mar 2015
72. Bateman C (2014) Chaos ethics. Zero Books, Winchester
73. Silvia PJ (2014) Interest – the curious emotion. Curr Dir Psychol Sci 17(1):57–60
74. Green MC, Brock TC, Kaufman GF (2006) Understanding media enjoyment: the role of transportation into narrative worlds. Commun Theory 14(4):311–327
75. Aristotle (335 BC) Poetics (trans: Butcher SH). Internet Classics Archive. Available online: http://classics.mit.edu/Aristotle/poetics.mb.txt. Accessed 9 Mar 2015.
76. Campbell J (1949) The hero with a thousand faces. Pantheon, New York
77. Vogler C (2007) The writer's journey: mythic structure for storytellers and screenwriters. Michael Wiese Productions, Studio City
78. Curry G, Ravenscroft I (2002) Recreative minds: imagination in philosophy and psychology. Oxford University Press, New York
79. Carroll N (1990) The philosophy of horror; or, paradoxes of the heart. Routledge, New York

80. Andrade EB, Cohen JB (2007) On the consumption of negative feelings. J Consum Res 34:283–300
81. Walton KL (1978) Fearing fictions. J Philos 75(1):5–27
82. Gaut B (2003) Reasons, emotions and fictions. In: Kieran M, McIver Lopes D (eds) Imagination, philosophy, and the arts. Routledge, London, pp 15–34
83. Bateman C (2014) Am I afraid of daleks? Quasi-fear as alief. Working paper. University of Bolton
84. Gendler TS (2008) Alief and belief. J Philos 105(10):634–663
85. Gendler TS (2008) Alief and belief in action (and reaction). Mind Lang 23(5):552–585
86. Markström AM, Halldén G (2009) Children's strategies for agency in preschool. Child Soc 23(2):112–122, March
87. Frasca G (2001) Rethinking agency and immersion: video games as a means of consciousness-raising. Digit Creat 12(3):167–174
88. Punday D (2005) Creative accounting: role-playing games, possible-world theory, and the agency of imagination. Poet Today 26(1):113–139
89. Harrell DF, Zhu J (2008) Agency play: dimensions of agency for interactive narrative design. In: Proceedings of the AAAI 2008 spring symposium on interactive narrative technologies II. Stanford, pp 156–162
90. Wardrip-Fruin N, Mateas M, Dow S, Sali S (2009) Agency reconsidered. Proceedings of DiGRA International Conference: Breaking New Ground: Innovation in Games, Play, Practice and Theory. Available online: http://www.digra.org/digital-library/publications/agency-reconsidered/. Accessed 10 Mar 2015
91. Tanenbaum K, Tanenbaum J (2010) Agency as commitment to meaning: communicative competence in games. Digital Creativity 21(1):11–17
92. Fredricksen E (2002) Progress quest [PC videoame], self-distributed
93. Midgley M (1974) The game game. Philosophy 49(189):231–253
94. Marcel AJ (2003) The sense of agency: awareness and ownership of action. In: Roessler J, Eilan N (eds) Agency and self-awareness. Oxford University Press, Oxford, pp 48–93
95. Sen A (1985) Well-being, agency and freedom: the Dewey lectures 1984. J Philos 82(4):169–221, April
96. Hribal JC (2007) Animals, agency, and class: writing the history of animals from below. Hum Ecol Rev 14(1):101–112
97. MacIntyre A (1999) Dependent rational animals: why human beings need the virtues. Gerald Duckworth and co., London
98. McFarland SE, Hediger R (2009) Animals and agency: an interdisciplinary exploration. Brill, Leiden
99. Persson M (2009) Minecraft [PC videogame], self-distributed
100. Tale of Tales (2005) The endless forest [PC videogame], self-distributed
101. Tale of Tales (2012) Bientôt L'été [PC videogame], self-distributed
102. Team Ico (2005) Shadow of the colossus [PS2 videogame]. Sony Computer Entertainment, Tokyo
103. Arika (2007) Endless ocean [Wii videogame]. Nintendo, Kyoto
104. Kant I (1790/2005) Critique of judgment (trans: Bernard JH). Dover Publications, Mineola
105. Newman J (2008) Playing with videogames, chapter 6. Routledge, London
106. Parker F (2008) The significance of jeep tag: on player-imposed rules in video games. Loading . . . , pp 1–3

Chapter 2
Affect Channel Model of Evaluation in the Context of Digital Games

J. Matias Kivikangas

Abstract Psychological emotion theories are underused in digital game research, possibly because they are divided into several competing camps and because they do not provide a framework easily applied to a context of digital games. I present a first step towards an integration of the camps, especially by combining Panksepp's view on primary processes, Scherer's component process model, and Cacioppo and others' evaluative space model. While specifying the different parts of the affect channel model of evaluation, I discuss how they are likely related to common game-related phenomena.

Introduction

Despite the wide recognition of the importance of emotions for game experience, the knowledge provided by **psychological emotion theories** has been little utilized in game experience research. One reason, no doubt, is the fragmented situation of the emotion theories: after a century of emotion research, different models attempting to explain how emotions work are counted in dozens, if not hundreds, and the researchers still cannot agree on what an emotion is (see the two Special sections on the topic that do not reach a conclusion, in the journal Emotion Review: [15, 37]). With the theories also often focusing on very specific features, they are also difficult to apply to a specialized field with a complex and still relatively poorly understood stimuli, such as digital games. As a result, most game researchers, who come from a wide range of backgrounds, have developed their own ideas on how emotions might contribute to the game experience that have little or no connection to the literature of emotion theories (for a rare exception, see [4, 20]). Mostly, only psychophysiological game research has been referring to emotion theories as a background (e.g., [25, 26, 34]; see [17], for a review of psychophysiological game studies).

J.M. Kivikangas (✉)
Department of Information and Service Economy, Aalto University, Helsinki, Finland
e-mail: matias.kivikangas@aalto.fi

Although individual theories do not give much answers, looking at them closely results in finding that many of them are not irreconcilable. In this chapter, I describe my interpretation of combining several emotion theories, and my suggestion how the combination can be used to explain phenomena related to game experience. Acknowledging the irony [28], I present my own model, the Affect Channel Model of Evaluation (ACME).

Background

While the theorists seem to be more interested in developing their theories in a rather limited area, the connections between different emotion theories are actually quite numerous when one knows where to look for them. Without the space to go into specifics here, in short, Panksepp's primary processes and LeDoux's survival circuits seem basically different descriptions of the same neural patterns (e.g., [21, 32]). Most emotion theories agree on some kind of appraisals that, in turn, have commonalities with the neuroscientific evidence (e.g., [5, 40]). In addition to obvious connection between core affect [35] and evaluative space (originally by [6]; see [29], for the current situation), the latter idea also overlaps LeDoux's "global organismic states" [21]. Thus, ACME is my synthesis: especially Panksepp's account on primary processes that form the neuroscientific base of specialized neural circuits adapted to specific evolutionary challenges; Scherer's Component Process Model (CPM) that provides the organization of laying the primary processes/survival circuits in a temporal order according to appraisals that activate them; and Cacioppo and others' Evaluative Space Model (ESM) that describes the global motivational state of the system that affects and is affected by appraisals and primary processes. The model is further influenced by Russell's [36] constructionist views that emphasize domain-generality[1] and higher conceptualizing processes that happen after the initial evaluations but that affect their next iterations by the various feedback loops.

ACME presents an interpretation of the automatic **evaluations** (or appraisals[2]) and the process cascades they activate non-consciously within about a second from the perceived change (cf. metaphorical "System 1" by [16]). Particularly, I attempt to find the time frame in which the evaluations might be processed, tied to the type of processing required for the evaluation to be possible (the processing levels inspired by four levels by [39]). I conceptualize the evaluations and the resulting neural activation spreading by describing **affect channels**, the evolution's way of organizing the contradictory action tendencies into coherent, prioritized response

---

[1]The "domain-generality" does not imply that the functions are not specialized—only that the domain of specialization is not "emotion" (cf. [19], Chapter 2).

[2]I prefer 'evaluation' over 'appraisal', because the latter is a strongly loaded term specifically related to appraisal theories and the theoretical constraints related to that literature.

modes. The present model is a gross simplification[3] of the intricate details provided by the abovementioned theorists, but it lays out the general structure of the system that produces affective responses.

As a clarification, *ACME is not primarily a model of emotions*, *but of motivational evaluations*. The name of the model reflects that: instead of referring to emotions, the 'affect' of affect channels refers to any feelings that move us (cf. [32]). What we call emotions simply happen to be among the most recognizable outputs of the evaluation processes.

Why not emotions? Any model of mind must be subordinate to the physical reality (described by neuroscience) and an evolutionary explanation of how it has evolved (e.g., [43]). There is no reason to posit the existence of a unitary "emotional system"—instead, evolution favored development that ended up, one by one, in a collection of processes that do their part in discerning certain survival- and procreation-relevant stimuli (evaluation) and preparing the organism for suitable action (motivation). Intertwined with other functions like perception, attention, memory, and so on, this covers all kinds of affective responses—such as pain, hunger, balance—without limiting to 'emotions' alone. I share the doubt Russell voiced [36], whether the term 'emotion' has any scientific value, with its referents so nebulous and arbitrary; like LeDoux [22], I prefer to use the word like a layperson would, as a non-scientific descriptor referring to those subjective feelings we consider emotional, including the landscape of feelings related to the game experience.

Model Details

Building Blocks

ACME posits that the evaluation system consists of the following parts: the low-level modules, affect channels, and the global evaluative state. In addition, I have organized the model according to two criteria: processing levels and biological priority order (Fig. 2.1).

The low-level **modules**[4] are functions that most likely are actual physical structures that can be—and in some cases already have been (e.g., so-called fear

---

[3]I have ignored, among many details, the whole system of homeostatic functions (pain, hunger, thirst, uncomfortable temperature, fatigue) which I contend should be recognized as an affect channel on their own right (cf. [31]).

[4]Following Kurzban [19], by "module" I mean "information-processing mechanism specialized to perform a particular function"—not the strong Fodorian module. Although I assume that the neural substrates of the modules can be found in the brain (see below), I do not assume that different modules are necessarily distinct from each other on the neural level. Therefore, I use expressions like "a module x is based on module y", meaning that the neurons that carry out the functions are largely the same, but because the modules are related to the function, the modules may be different.
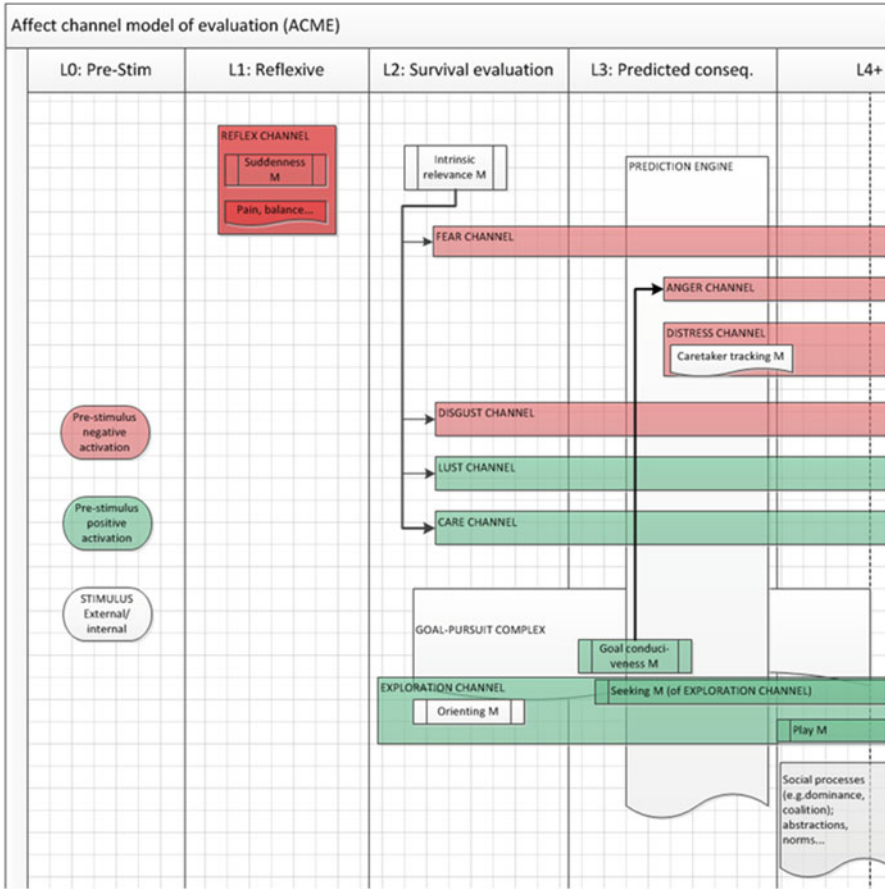
Fig. 2.1 Affect channel model of evaluation. Ovals in L0 represent baseline effects. *Boxes* represent affect channels or modules; *red color* indicates a primarily negative influence on GMS, *green* a primarily positive influence. All channels are activated by evaluative modules, but only the known modules are shown. *Wavy boxes* indicate that something like that should exist, but the details for including them in the model are unclear

circuit: [23])—found in the brain. I sometimes discuss the two functions of these modules, the evaluator and the mobilizer, as if they were separate, because they are based on ideas from different theories (appraisals and primary processes, respectively). On the neural level, however, they are most probably simply two functions of the same module, as modules are evolved as adaptations for a particular function, and it does not seem likely that these functions would have evolved separately.

The evaluators, based on stimulus evaluation checks by Scherer [39, 40], process the sensory information[5] according to a particular adaptively relevant question, regardless of whether it originates from the sensory organs or is internally created by recall or imagination; in a sense, their output is an answer to one question (such as "is this stimulus new?"). Evaluators determine which mobilizer should be started, if any, from an evolutionarily predetermined set. When activated, the mobilizers, based on primary processes [32] and survival circuits [21], launch (again, evolutionarily predetermined) large-scale activation changes both in the brain and in the autonomic and somatic nervous systems (ANS and SoNS), resulting—if not inhibited by other processes—in observable changes in emotion components. I list modules for only some channels mainly following the appraisal theories, but obviously all channels have some kind of modules for evaluating the stimulus—we just lack the empirical details.

The **global motivational state** (**GMS**), based on the evaluative space model [29], represents the extent the organism is in positive (approach) and/or negative (avoid) motivational state. The GMS is changed by most mobilizers, and it affects all neural modules, positivity inhibiting negative evaluations and activation spread while facilitating positive evaluations and activation, and negativity influencing evaluations and activation in the opposite manner (cf. mood congruency; [35], p. 156). Although the effects of an evaluative space are well supported by evidence (e.g., [30]), it is unclear how the GMS is manifested on the physical level (i.e. where is the information on positivity and negativity stored?). For the purposes of this chapter, the evidence supports the treatment of the GMS as an abstract "state" omnipresent for neural modules.

Affect channels, therefore, are the specific patterns of neural modules that follow a particular adaptive function in order to produce suitable behavior. On the level of affect channels I also separately postulate two important **general-purpose functions** (goal-pursuit complex and prediction engine) that are not primarily evolved to produce behavior but to help affect channels to do their jobs. I have inferred their existence by the functions required by the model, but their neuroscientific basis is currently unclear. The individual affect channels are briefly described after explaining how the model is organized.

---

[5]Note that by "stimulus" I do not mean, for example, a single seen object, as the visual system makes the distinction between objects relatively late ([10], Chapter 5). Instead, I mean the information that reaches an evaluator after being processed by different perception modules into some format that it can evaluate. That is, like the perception modules, the evaluators focus on very specific features of the perceptual information, from a simple feature like darkness, to a highly processed understanding of the environment and context. This also means that while I assume an external stimulus (such as a digital game), the model does not ignore self-caused (imagined or recalled) stimuli, that are treated no different from the external stimuli when they are fed to the evaluators. This is supported by the vast evidence that imagined situations lead to same kind of physical changes in the brain than external stimuli (e.g., [13]).

Organization

To promote adaptive behavior despite conflicting motivations between the different mobilizers, the affect channels must have a biologically wired **priority order**. This is the sensitivity to activation of the channels more relevant for imminent survival, compared to those less relevant, resulting in an ability to easily override motivations of less pressing concern. The priority order does not imply fixedness. Rather, it is the relative strength between the channels—a channel higher in the order needs less activation to override a channel of lower priority. Conversely, a sufficiently strong activation on a lower-priority channel may still result in the individual engaging in behavior where the mild concerns for survival are temporarily suppressed. The GMS further tilts the table so that certain evaluations and output activations are more likely than others, making overriding easier (negative activation) or more difficult (positive activation).

To provide further structure between the channels and their contents, ACME is organized according to the **processing levels** respective to the complexity each neural module requires. (These levels are meant as a tool for understanding the relationships, not as descriptions of natural categories.) Simultaneously, the levels roughly describe the relative time frame in which its processes finish[6]—although likely all survival-related processes launch as early as possible, the more complex processing goes through more complex neural networks and therefore takes more time. The levels are: pre-stimulus level (L0), reflexes (L1), survival evaluation (L2), evaluation of predicted consequences (L3), and complex, conceptual evaluation (L4+) processes. As the L0 implies, the time frame is relative to a moment when a new stimulus is detected. Although Cunningham and others' [9] critique—that there is no "time zero" because all processes are running all the time and the previous activity acts as a powerful biasing factor for further processing—is valid, the greatest changes occur when a new stimulus is detected, making it the best reference point.

Like the whole CNS, the modules are connected in a heterarchical way [29], many of them running in parallel and some earlier processes serving as the necessary activators of some later, but processes also being activated, facilitated, and inhibited by processes from higher levels. For example, most of the early processes, although powerful in directing the organism for action, can be inhibited by conscious effort. The processes are recursive and are updated constantly, but I mostly discuss only the first iterations.

Further, it is important that while the late processes may affect the next iterations of the early processes by inhibiting (or facilitating) the outputs of the mobilizers, the evaluators activating the mobilizers is involuntary (further supporting the assumption that actually the two are two parts of the same module). This is apparent in situations like flinching when a sudden movement is detected near the head, feeling scared when alone in the dark, or feeling lust in presence of strong sexual

---

[6]I also note that the levels seem to correspond roughly to the (probable) evolutionary and developmental order of appearance, but this is not the main purpose of the levels.

cues. If the evaluation is made, some kind of response (although it can be dampened) is inevitable. The only way to avoid the response completely is to change the situation so that the evaluation is never made.

Finally, conscious awareness is not assumed to occur clearly at one point in this time-frame—instead, I believe it is gradually constructed from several different processes within a longer time window, and it only gains access to a small part of information from the processes described here (cf. [19]).

Affect Channels

Some of the affect channels can be further labeled based on their urgency related to survival. Those that handle evaluations requiring urgent responding are higher in priority than those that evaluate stimuli with non-urgent responses. Urgency does not equate priority, however: for example, the Anger channel is not evolved for survival-urgent situations, but its priority is high.

The evolutionary principle of conserving resources dictates that if it is not necessary to act, the organism is better off resting than spending valuable resources. On the other hand, when the resources are not scarce so that they should be saved only for the necessities, it is useful that the organism ensures the future survival by securing more resources and learning the environment. The **Exploration channel** is the implementation of this function. Its activation is of the lowest priority: it directs behavior most when nothing is evaluated to be threatening the organism, no particular goals are pursued, and the resource-gathering modules are inactive. When nothing more pressing is requiring attention, it drives the organism to explore and find relevant resources. Originally that has meant food, water, shelter, and mate, but other adaptations have expanded the domain of this channel to social and abstract resources (such as knowledge) as well.

Although the **Goal-pursuit complex**, on the neural level, is likely an extension for the Exploration channel added with currently unknown other components, I discuss it separately because it clearly forms a general-purpose function. It gives the other channels the tools to pursue specific goals, rewarding the seeking—not only of new things, but of anticipated things. The other channels likely also have separately their own each reward mechanisms, producing positive feelings in the end of their respective action mode (sexual gratification is the clearest example, but, e.g., the removal of anger-inducing obstruction is also satisfying).

The non-urgent survival channels are essential for survival, but not so urgent that something must be done about them immediately. After a resource is found, it should be consumed. In the case of food and water this is simple: separate consumption modules (not presented in ACME) kick in that reward drinking and eating. In case of mating the process is more complex, and requires a broader response pattern to ready the body and to signal that readiness, implemented by the **Lust channel**. Because the channels are evolutionarily adapted, the organisms do not only secure their own well-being, but also that of their offspring. The

Care channel works in tight interaction with the Exploration channel, promoting resource-securing behavior for the offspring as well, as well as looking after and keeping the young themselves in safety.

The activation of the urgent survival channels requires immediate action. The **Disgust channel** is the least pressing, driving the organism to avoid potential sources of unhealthiness—carcasses, bodily wastes, and spoiled foodstuffs, but also outsiders that carry a greater risk of disease for the whole community [14]. The **Distress channel**, especially active in the young and the nurturing mothers, keeps track of the mother/offspring and promotes, again in tight interaction with the Goal-pursuit complex and the Exploration channel, behavior to reunite lost loved ones. The **Fear channel** reacts to the myriad signals of imminent threat in the environment and has an elaborate array of evaluation and mobilization circuits at its disposal to advance survival.

The **Anger channel** is probably located between Distress and Fear in respect to priority, although its urgency for survival is not typically high. This channel responds when the goals set for the Exploration channel to pursue are obstructed, mobilizing resources to remove the obstruction by force.

Finally, the **Reflex channel** responds immediately and automatically to direct damage, or the threat of it, by moving the body out from the immediate danger. Its priority and urgency are the highest, but the responses are also so quick that it does not hinder other survival channels activation much.

The Model

Pre-stimulus Level (L0)

The **pre-stimulus level** describes the processes that can be assumed to be running in the absence of any particular stimulus that would have been evaluated to require a response. This level does not imply anything about the time frame of its processes, since they are not launched by detecting a new stimulus.

Exploration Channel In absence of other activation (and of scarcity), the **seeking module** constantly makes the organism look for something rewarding, instead of doing something that does not give immediate rewards (i.e., procrastinating, channel surfing with TV, constantly checking social media). In the brain, this is driven by the dopamine system which responds to novel, attention-grabbing events, but stops responding when the stimulus grows too predictable ([21, 32]; see also: positivity offset, in [29]). The module evaluates the novelty of the stimulus (cf. novelty check in [40]), and directs attention to those evaluated new while increasing action readiness (arousal). Interacting with higher processes (see L4+, below), it also evaluates stimuli that carry the possibility of finding something new—process that is expressed as mild interest. In games, this results in the eponymous exploration behavior in sandbox worlds (is there something interesting behind those hills?)

and continuous completion of easy tasks when it carries the promise of a reward (clicking away in Candy Crush, grinding in World of Warcraft, or the "one more turn" effect in Civilization and its kin; see [18], for a similar idea of play as exploration). The activity itself is not that interesting and the rewards are not actually that satisfying when they are obtained, but the design that has always something new behind the corner activates the seeking module which continues until some other process stops it—or when the activity grows so predictable that it does not produce new rewards anymore.

Other Channels Of course, the "absence of other activation" is not a trivial criterion. Evolutionarily, seeking new resources is useful only when the imminent survival is not threatened. Contemporary humans rarely have to be afraid of predators, but hunger and very hot weather turn the seeking module into finding relief instead of exploration. Similarly, stress, worry, and irritation prevent the seeking module to kick in. In depression, nothing feels like anything anymore— the seeking module does not work, does not give reward for finding new things, leading to apathy. Positive feelings can prevent exploration as well: the activation of Lust channel sets the goal to sexual gratification, foregoing leisurely exploration in order to seek sex ([32], Chapter 3)

Reflexes (L1)

The first level describes processes that occur reflexively: activation of SoNS is launched immediately when a module calls for it, directly from the subcortical brain regions without waiting for further higher-order processing. This may happen before 100 ms from the stimulus onset.

Reflex Channel The very rudimentary features of the stimulus are evaluated by the **suddenness module** (cf. suddenness subcheck in [39]), detecting sudden loud and abrupt sound or quick and large movement in the visual field. This activates the startle response, resulting in an increased alertness (guiding attention to scan the perimeter instead of focusing intensely to one target) and elevated action readiness (heart pounding and palms sweating; increased GMS negativity) in response to moderate activation, and in addition to these, dodging and shielding movements when the activation is high. All this happens nonconsciously, and the information about their occurrence reaches the consciousness only afterwards.

The suddenness evaluation is exacerbated by pre-existing negative and slightly inhibited by positive GMS activation. The higher-order processes of anticipation also inhibit the response, to an extent, as the organism's understanding of the environment creates expectations: if the organism readily expects a loud sudden noise (e.g., a bang of the player's own weapon or when a previously detected monster crashes through the window), the startle response will not be activated by

it. However, this only applies to the expected stimuli (i.e., if a bang was expected, a sudden "boo!" from a person behind the player still gives the response).

Pain and balance reflexes can probably be included in the Reflex channel as well, but they do not seem relevant in gaming context.

Survival Evaluation (L2)

The second level evaluations are processed a bit further, **pattern**-**matching** the stimulus information to genetically inherited patterns, recently detected stimuli, or the current low-level concerns set by higher processes. The patterns are still relatively simple because the responses need to occur quickly, around 100–200 ms from the stimulus onset [12]: they are adaptations that protect from immediate threats relevant to our hunter-gatherer ancestors.

Survival Channels The main survival evaluation is common to most survival channels. The **intrinsic relevance module** compares the current stimulus features to evolutionarily relevant patterns that signal relevance to survival [40]. Darkness and stimuli that resemble spiders, snakes, or angry and violent faces activate the Fear channel, and stimuli resembling human waste or other disease carriers activate the Disgust channel, both resulting in withdrawal motivation and higher action readiness. The detection of sexual cues more the Lust channel, and detection of nurturance cues the Care channel. When the GMS is already negatively activated, the module evaluates things easier as threatening, and when positively, the sexual and nurturance cues are evaluated as stronger and more likely.

In games, the primal fear cues, and to some extent disgust cues as well, are commonly used in environment and enemy design, because they fire up arousal and activate suitable threat-related associations, creating a suspenseful mood, and they mobilize action that lead to gratification when the goal has been reached (i.e., removal of threatening entity). While in a virtual world the Lust channel cannot (at least currently) reach its goal, sexual cues are often used for their arousal-inducing effects (although mostly for heterosexual males). In games that require looking after some characters, the game designers often use nurturance cues such as big eyes and soft and round facial features resembling infants, because they activate warm and fuzzy feelings and care tendencies—if you go "aww" upon seeing something, the intrinsic relevance module has activated the Care channel.

Exploration Channel When the stimulus is not intense enough to elicit a startle reaction but is evaluated as novel, the **orienting module** elicits an orienting response. It probably uses at least partly the same circuitries as the more complex seeking module, reacting to novelty, and might be linked to the intrinsic relevance module as well. Orienting response is a basic tool in the human attention system but does not have much significance for games per se.

In games, all goal-directed action involves assessing whether a particular stimulus is relevant for the current goals or not. According to the empirical evidence (the "concern pertinence" check in [12]), the early modules of the **Goal-pursuit complex** evaluate the stimulus for goal relevance in this time window. Considering that the other evaluations on this level probably only do simple pattern-matching, the processing here are not likely to be much more complicated. As the processing behind goal-pursuit in general however are necessarily more complicated than that (evaluating, e.g., whether a certain change in the zerg movement patterns in StarCraft 2 requires a response or not in respect to the player's goals), the early modules are likely provided preprocessed patterns to match (e.g., react when those mutalisks start moving).

Evaluation of Predicted Consequences (L3)

The next level of processing complexity goes beyond simple pattern-matching to involve associative memory, and specifically the other general-purpose function: the **Prediction engine**. In brief, I use the term to refer to the function where the stimulus is associated with similar occurrences in memory to see what happened on those previous situations, and therefore to predict what might happen now. The initial, immediate prediction is the most available situation, which might or might not be the most likely one (depending on how the neural weights have been arranged earlier).

The process is still very much automatic and nonconscious—relevant processes occurring around 300–600 ms[7] [11]—and it probably uses innate logical structures that, for instance, deduce causality from sequential events (named "causal attribution" or "agency" by appraisal theorists: [41]; attribution also discussed by [35]), as famously discussed by David Hume. Until the capability to attribute causality to perceived agents or to predict consequences of events, the stimuli are without any social meaning, and for example the fear response cannot be more complex than to turn around and run. With this first bit of contextual information, much better goals can be set for the Goal-pursuit complex to seek.

Goal-Pursuit Complex and Anger Channel The most obvious result of the interaction with Prediction engine is the evaluation of obstruction or furthering the current goals by the **goal conduciveness module** (or appraisal: [38]). If the

---

[7]How can the elite Counter-Strike players play this extremely fast-paced game, if the simple evaluation of the consequences of an action is supposed to take half a second? The experiments about the timings have been carried out with abstract tasks that people have no previous experience with. When a familiar environment is navigated, the typical situations and their consequences are already associated to their evaluations and behavioral responses, allowing quick responding by automated motor patterns that border reflexes in extreme cases. Instead of absolute timings, the levels are meant to indicate the relative processing speeds of different processing types—the milliseconds are secondary.

goal conduciveness module evaluates the stimulus event to further the goals, it increases positive GMS. However, obstruction of goals activates the Anger channel, mobilizing bodily resources for removing the obstruction—and ultimately leading to the subjective feeling, the distinctive flare of anger or frustration. In gaming, this evaluation occurs, for instance, when the computer freezes just as you were doing something, or when other people get in your way.

Distress Channel When the track of the caretaker or offspring (or probably any other strongly bonded individual) is lost by a caretaker tracking module of some kind (appraisal theories do not provide a clear indication of what this might be), the Distress channel is activated ([32], Chapter 9), causing social signaling for needed help and (for the caretaker) mobilizing resources for getting the offspring back by force if necessary. If a permanent or long-time separation is predicted, the behavioral components are inhibited but the activation still remains, manifesting as sadness and grief. Games are almost never able to create enough bonding to virtual characters to make use of this channel, although exceptions have recently appeared (e.g., Telltale's Walking Dead).

Other channels The seeking module on the Exploration channel, enabled by the Prediction engine, was already described in L0. The predictions also affect the activation on other channels. As mentioned, with better understanding of the context, the Fear channel and the Goal-pursuit complex can now set more adaptive goals, but similarly, the responses by Disgust, Lust, and Care channels are modified by the new information. For example, it is possible that the inhibition of primal behavior occurs at this level: you don't habitually avoid the plastic fake vomit you find in your prank box, or you simply nod while listening to the extremely attractive person next to you, because the module in your head predicts that those approaches do not result in undesired consequences. When the activation is not extremely strong (or the learned regulation mechanisms particularly weak), adult humans can and do inhibit most of their responses that the evaluative system promotes. In a gaming context, inhibiting the Fear channel behavioral responses—while still getting the arousal activation from the evaluations—is the basis of the horror games.

Conceptual Evaluations and Further (L4+)

Many emotion theorists say that later in the construction of emotion "conceptualization" and "categorization" occur [2, 36], but unfortunately, there is little empirical information on what the more specific processes are or even what would be their general structure. One step before moving into full-blown abstract thinking is the broader contextualization of stimuli, utilizing more complicated predictions and the comprehension of more complex relationships between different agents, events, situations, objects, and time concepts, while abstractions of these relationships into general rules seem to be another step. However, with this vague understanding it is difficult to identify clear modules or the principles on which we might make

distinctions between their processing levels. Scherer (e.g., [40]) and other appraisal theorists, for example, specify norm compatibility as a separate appraisal, but it is not clear what the process actually is (is there one process or many? what kind of requirements they might have?). Most of them are also out of the focus of this paper, so apart from one exception below, I leave the higher processes on levels 4+ untouched.

For the game context, the most interesting module can be inferred from Panksepp's work ([32], Chapter 10): along with the capabilities for understanding social context, the **play module**, strongly based on the seeking module, is enabled. The evolutionary function of the play module is assumed to be a way for the organism to learn physical and social rules and skills[8] (i.e., seeking informational and social resources) by testing them and their limits in a safe way. The example of this is a rough-and-tumble play that occurs with all mammals, where the young test the relationships they have with each other and with adults (e.g., an adult or a pup that have been established as stronger will pretend losing to a weaker one; [32], Chapter 10). This kind of a second-order seeking is only possible when the young begin to understand the social world as a new environment to explore: enough to being able to predict it somewhat (seeking does not activate in the first place until the prediction engine can predict finding something), but not too much (activation of the seeking module stops when nothing new can be found). However, the play of this kind is still first and foremost *activity*—the behavior patterns the module produces.

Interactions between the play module and other processes expand the variety of rules and skills that are explored. For example, interaction between the play module and the Lust affect channel has an important role in finding and courting a mate in many animals, although the complexity of the social world of humans makes it more difficult for us. Competition, in turn, is arguably a result of interaction with the (probably higher-order) processes related to social dominance. With the development of more abstract understanding, humans can extend the exploration to abstract rules and systems as well, especially in sports. Digital games engage (mostly) this kind of exploration with heavily automatized play in eSports, but also with game design/player behavior patterns like leveling up (i.e., finding out new ways to do things) and mastering the behavior of the game (such as movements of the character in the game world; i.e., honing the skills and the predictions like animals in a rough-and-tumble play).

---

[8]Note that albeit the evolutionary function of the play module is to learn survival-relevant skills and understanding, the modules themselves do not regulate their operation according to whether there is actual use for the skills and understanding in the life in general. Procrastination and diabetes occur for the same reason: the modules behind our behavior have evolved in a much more resource-scarce environment where *too much resources* was never something we needed to adapt to. Play is the same—it was never "meant" to be available as often as it is in the modern world where we don't have to focus on survival all the time. All play the individual had time for in the ancestral past was a bonus.

In humans, a still more complex form of exploration—a third-order seeking following the same principle of "predictable but not too predictable"—can be seen in the what-if, pretend, or role play: the exploration of relationships and social rules *that do not exist* between the players outside the game[9] (as opposed to testing the existing relationships as in second-order seeking). This requires a further capability to understand and imagine counterfactual relationship networks (i.e., fiction) and to accommodate one's own behavior according to them. I assume that pretend play is not a result of a separate adaptive module like play likely is, but instead a product of the interaction between the play module and higher-order processes (that are responsible for, e.g., the theory of mind; [24]) that feed activation to it.

As opposed to second-order seeking, what-if play is more *imagining* than activity: the seeking module can now be activated by (and therefore provides satisfaction for) mental simulation of abstract rules (e.g., chess, Magic: the Gathering), but also fictional scenarios. This is first introduced in activities such as storytelling and children's role play,[10] but occurs later for instance in fan culture, role-playing games, and sexual fantasizing. In game context, it creates an additional layer over games that have a fictional theme (cf. ameritrash vs. eurogame board game designs). A further special mode of what-if play is enabled with higher-order processing of regulating one's own affective responses to a degree by contextualizing the situations that the primal mechanisms in the affect channels evaluate. By contextualizing a situation that activates negative affect channels (Fear, Distress, or Disgust) as safe, the rewards granted by the Exploration channel modules enhanced by increased arousal can be enjoyed (e.g., horror or tragedy; [1, 3]). Many forms of gaming also exhibit this dynamic, such as extreme role-playing [27]. Of digital games, the horror genre is an obvious example.

Final Words

The current model is based on theory reading and simplified considerably. It is not empirically tested yet, although the theories and empirical works it is based on have notable support (see section "Background"). Furthermore, my own expertise is in psychophysiology and emotion psychology, so while neuroscience is an important part of the current work, I concede that I am not an expert of that field. Given time, hopefully the inevitable mistakes and misunderstandings will be corrected. Nevertheless, the core of the model is the integration of evidence and theories created by others, not in the details I have conjectured. With the final note below, I invite critics to expose the mistakes, in order to see whether the core can stand when the details are corrected.

---

[9]Cf. magic circle: [42].

[10]Not in early pretend play, though: young children apparently mimic the activities, without imagining the mental worlds of mom and dad or police and robber [24].

Most of the engagement, immersion, flow, etc. literature seems to describe something resulting from the Exploration channel and Goal-pursuit complex activation (see [7], for a review; see also [33]). Flow, for example, could be understood as a strong activation and interaction of the seeking/play module and the Goal-pursuit complex, when the activity provides clear and frequent milestones, is automated enough to utilize highly specialized motor patterns, and is challenging enough to keep results unpredictable. In addition to the empirical knowledge on the channels (or rather, the primary processes they are based on; [32]) and the timing of specific evaluations related to them [11, 12], there is practical information about the optimal arrangement of reward frequencies and probabilities, most likely indicating something about the brain processes behind their utilization, to be gained from the player data of contemporary massive games (see e.g., [8]).

References

1. Andrade EB, Cohen JB (2007) On the consumption of negative feelings. J Consum Res 34(3):283–300. doi:10.1086/519498
2. Barrett LF (2013) Psychological construction: the Darwinian approach to the science of emotion. Emot Rev 5(4):379–389. doi:10.1177/1754073913489753
3. Bartsch A, Vorderer P, Mangold R, Viehoff R (2008) Appraisal of emotions in media use: toward a process model of meta-emotion and emotion regulation. Media Psychol 11(1):7–27. doi:10.1080/15213260701813447
4. Bateman C, Nacke LE (2010) The neurobiology of play. In: Proceedings of the international academic conference on the future of game design and technology pp 1–8. ACM
5. Brosch T, Sander D (2013) Comment: the appraising brain: towards a neuro-cognitive model of appraisal processes in emotion. Emot Rev 5(2):163–168. doi:10.1177/1754073912468298
6. Cacioppo JT, Berntson GG (1994) Relationship between attitudes and evaluative space: a critical review, with emphasis on the separability of positive and negative substrates. Psychol Bull 115(3):401–423. doi:10.1037/0033-2909.115.3.401
7. Caroux L, Isbister K, Le Bigot L, Vibert N (2015) Player–video game interaction: a systematic review of current concepts. Comput Human Behav 48:366–381. http://doi.org/10.1016/j.chb.2015.01.066
8. Chatfield T (2010) 7 ways games reward the brain [video]. TEDGlobal. Retrieved from http://www.ted.com/talks/tom_chatfield_7_ways_games_reward_the_brain#t-966410
9. Cunningham WA, Dunfield KA, Stillman PE (2013) Emotional states from affective dynamics. Emot Rev 5(4):344–355. doi:10.1177/1754073913489749
10. Gazzaniga MS, Ivry RB, Mangun GR (1998) Cognitive neuroscience. The biology of the mind. W.W. Norton & Company, New York
11. Gentsch K, Grandjean D, Scherer KR (2013) Temporal dynamics of event-related potentials related to goal conduciveness and power appraisals. Psychophysiology 50(10):1010–1022. doi:10.1111/psyp.12079
12. Grandjean D, Scherer KR (2008) Unpacking the cognitive architecture of emotion processes. Emotion 8(3):341–351. doi:10.1037/1528-3542.8.3.341
13. Holmes EA, Mathews A (2010) Mental imagery in emotion and emotional disorders. Clin Psychol Rev. doi:10.1016/j.cpr.2010.01.001
14. Horberg EJ, Oveis C, Keltner D, Cohen AB (2009) Disgust and the moralization of purity. J Pers Soc Psychol 97(6):963–976. doi:10.1037/a0017423
15. Izard CE (ed) (2010) On defining emotion [Special section]. Emot Rev 2(4):363–385
16. Kahneman D (2011) Thinking, fast and slow. Book. Farrar, Straus and Giroux, New York

17. Kivikangas JM, Chanel G, Cowley B, Ekman I, Salminen M, Järvelä S, Ravaja N (2011) A review of the use of psychophysiological methods in game research. J Gaming Virtual Worlds 3(3):181–199. doi:10.1386/jgvw.3.3.181_1

18. Koster R (2013) Theory of fun for game design. O'Reilly Media, Inc., Sebastopol

19. Kurzban R (2010) Why everyone (else) is a hypocrite. Evolution and the modular mind. Princeton University Press, Princeton

20. Lang A (2006) Motivated cognition (LC4MP): the influence of appetitive and aversive activation on the processing of video games. In: Messaris P, Humphreys L (eds) Digital media: transformation in human communication. Peter Lang Publishing, New York, pp 237–256

21. LeDoux JE (2012) Rethinking the emotional brain. Neuron 73(5):653–676. doi:10.1016/j.neuron.2012.02.004

22. LeDoux JE (2014) Comment: what's basic about the brain mechanisms of emotion? Emot Rev 6(4):318–320. doi:10.1177/1754073914534506

23. LeDoux JE, Phelps EA (2008) Emotional networks in the brain. In: Lewis M, Haviland-Jones JM, Barrett LF (eds) Handbook of emotions, 3rd edn. Guilford Press, New York, pp 159–179

24. Lillard A (1993) Pretend play skills and the child's theory of mind. Child Dev. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8624.1993.tb02914.x/full

25. Mandryk RL, Atkins M (2007) A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. Int J Hum-Comput Stud 65(4):329–347. doi:10.1016/j.ijhcs.2006.11.011

26. Martínez H, Yannakakis GN (2011) Analysing the relevance of experience partitions to the prediction of players' self-reports of affect. In: Affective computing and intelligent interaction, part II, LNCS 6975, Springer, Berlin/Heidelberg, pp 538–546

27. Montola M (2011) The painful art of extreme role-playing. J Gaming Virtual Worlds 3(3): 219–237

28. Munroe R (2011) Standards. Retrieved from http://xkcd.com/927/

29. Norman GJ, Norris CJ, Gollan J, Ito TA, Hawkley LC, Larsen JT, ... Berntson GG (2011) Current emotion research in psychophysiology: the neurobiology of evaluative bivalence. Emot Rev 3(3):349–359. doi:10.1177/1754073911402403

30. Norris CJ, Gollan J, Berntson G, Cacioppo JT (2010) The current status of research on the structure of evaluative space. Biol Psychol 84(3):422–436

31. Panerai AE (2011) Pain emotion and homeostasis. Neurol Sci: Off J Ital Neurol Soc Ital Soc Clin Neurophysiol 32(Suppl 1):S27–S29. doi:10.1007/s10072-011-0540-5

32. Panksepp J, Biven L (2012) The archaeology of mind: neuroevolutionary origins of human emotions (Norton series on interpersonal neurobiology). W.W. Norton & Company, New York

33. Peifer C (2012) Psychophysiological correlates of flow-experience. Adv Flow Res:139–164. Retrieved from http://link.springer.com/10.1007/978-1-4614-2359-1_8

34. Poels K, van den Hoogen W, IJsselsteijn WA, de Kort YAW (2012) Pleasure to play, arousal to stay: the effect of player emotions on digital game preferences and playing time. Cyberpsychol Behav Soc Netw 15(1):1–6. doi:10.1089/cyber.2010.0040

35. Russell JA (2003) Core affect and the psychological construction of emotion. Psychol Rev 110(1):145–172. doi:10.1037/0033-295X.110.1.145

36. Russell JA (2009) Emotion, core affect, and psychological construction. Cognit Emot 23(7):1259–1283. doi:10.1080/02699930902809375

37. Russell JA (ed) (2012) On defining emotion [Special section]. Emot Rev 4(4):337–393

38. Scherer KR (2001) Appraisal considered as a process of multi-level sequential checking. In: Scherer KR, Schorr A, Johnstone T (eds) Appraisal processes in emotion: theory, methods, research. Oxford University Press, New York, pp 92–120

39. Scherer KR (2009) The dynamic architecture of emotion: evidence for the component process model. Cognit Emot 23(7):1307–1351. doi:10.1080/02699930902928969

40. Scherer KR (2013) The nature and dynamics of relevance and valence appraisals: theoretical advances and recent evidence. Emot Rev 5(2):150–162. doi:10.1177/1754073912468166

41. Scherer KR, Schorr A, Johnstone T (2001) In: Scherer KR, Schorr A, Johnstone T (eds) Appraisal processes in emotion: theory, methods, research. Oxford University Press, Canary

42. Stenros J (2014) In defence of a magic circle: the social, mental and cultural boundaries of play. Trans Digit Games Res Assoc 1(2):147–185
43. Tooby J, Cosmides L (2008) The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In: Lewis M, Haviland-Jones JM, Barrett LF (eds) Handbook of emotions, 3rd edn. Guilford Press, New York, pp 114–137

Chapter 3
Affective Involvement in Digital Games

Gordon Calleja, Laura Herrewijn, and Karolien Poels

Abstract The chapter takes as its main object of study (Calleja's (2011) In-game: from immersion to incorporation. MIT Press, Cambridge, MA) Player Involvement Model, an analytical framework designed to understand player experience developed through qualitative research. The model identifies six dimensions of involvement in digital games, namely kinesthetic involvement, spatial involvement, shared involvement, narrative involvement, ludic involvement and affective involvement. The goal of the chapter was to develop the model further by testing it in an experimental context. Consequently, three experiments were conducted in order to examine how different components of digital gameplay (i.e. the story that is written into a game, the social setting in which a game is played, and the player's control in a game environment) can affect the player's involvement on the six proposed dimensions. Special attention is paid to affective involvement, and how this dimension of player involvement relates to the other dimensions. The findings of the experiments provide initial support for (Calleja's (2011) In-game: from immersion to incorporation. MIT Press, Cambridge, MA) Player Involvement Model in a quantitative setting.

Introduction

This chapter uses Calleja's [3] Player Involvement Model to organize an analysis of affect in games. The *Player Involvement Model* dissects player involvement with games, identifying six main dimensions of involvement (i.e. kinesthetic, spatial, shared, narrative, ludic and affective involvement) plotted on two temporal

G. Calleja (✉)
Institute of Digital Games, University of Malta, Msida, Malta
e-mail: calleja@itu.dk

L. Herrewijn
Department of Communication Sciences, Research Group CEPEC, Ghent University, Ghent, Belgium

K. Poels
Department of Communication Studies, Research Group MIOS, University of Antwerp, Antwerp, Belgium

phases: the immediate moment of game-play (i.e. micro-involvement) and off-line engagement with games (i.e. macro-involvement). Since the model was developed through qualitative research, we decided that the next step in its evolution would be to develop it further by quantitatively testing it in an *experimental set-up*. Consequently, we conducted three experiments in order to elaborately examine how different components of digital gameplay (i.e. game story, social setting, game control) can affect players' involvement on the micro level, and how the different dimensions of player involvement relate to each other. As this collection focuses on emotion and affect in games, the current chapter especially investigates the relationship between *affective involvement* and the other five dimensions of the Player Involvement Model.

In order to arrive at the investigation of affective involvement in combination with the other dimensions, we will first give a brief description of the different layers of player experience. Next, we will give an overview of the Player Involvement Model, outlining its six dimensions and two temporal phases. In the empirical portion of the chapter, we will then investigate how varying a specific component of a digital game (e.g. its story) affects players' involvement on all of the dimensions, and how these dimensions relate to and combine with each other.

The Bottom-Up Experience Triangle

One of the challenges in discussing player experience, at least within Game Studies, is the lack of differentiating between different forms of engagement with the game [3]. Authors often use terms such as engagement, involvement, attention, absorption, and sometimes even immersion, interchangeably (e.g. [2, 6, 7, 10, 15]). This makes it challenging to know what aspect of experience each author is actually referring to. In order to clarify this issue, we start from a layered model of player experience, adopted from the bottom-up experience model in cognitive psychology (see Fig. 3.1).

Beginning from the bottom and working our way up, each layer of the triangle acts as a pre-requisite for those that follow. We cannot, for example, experience involvement in a game without first paying attention to it. The first and most basic layer is wakefulness. *Wakefulness* is the basic state of being conscious. This is a biological, not a cognitive function. If we are unconscious, we obviously cannot be playing. Next, we have attention. *Attention* is the willed or automatic direction of our awareness to certain stimuli in the environment. Whereas attention deals solely with cognitive functions, involvement deals with the nature and quality of the thing we are directing our attention to. *Involvement* considers the emotional spectrum and thus the whole, embodied experience. *Incorporation* [3] refers to the experience of inhabiting the game environment, sometimes referred to as "presence" or "immersion". This phenomenon occurs as a result of the blending and internalization of involvement and is thus the most elusive of the layers as it is experienced at a subconscious level.
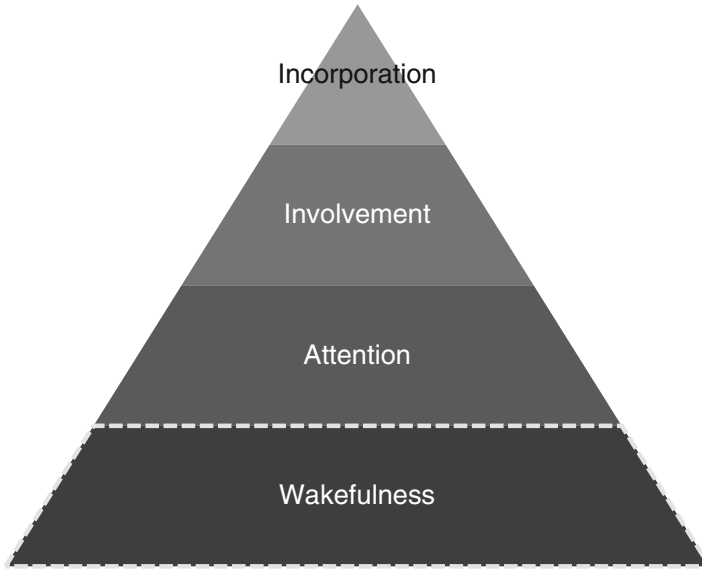
Fig. 3.1   The bottom-up experience model of player experience

This chapter will deal primarily with the *involvement layer of the experience triangle*. While attention will also play a part in our discussion of affect, its utility is primarily in describing the structure of the cognitive resources that we have at our disposal and how we direct these resources during game-play. It is involvement that describes the quality of the game-playing experience and thus its affective dimension. The rest of the chapter will therefore describe the role of affect within player involvement in greater detail. First of all, however, we will give a brief overview of involvement as a whole.

The Player Involvement Model

In his book-length treatment of player involvement, Calleja [3] uses his model as a foundation upon which to build further investigations into player experience. His Player Involvement Model identifies *six dimensions of involvement in digital games* (i.e. kinesthetic, spatial, shared, narrative, ludic and affective involvement), each considered relative to *two temporal phases* (i.e. the macro-involvement phase and the micro-involvement phase) (see Fig. 3.2).

The six dimensions concern involvement related to (1) control and movement in the game environment (*kinesthetic involvement*), (2) the exploration, navigation and learning of the game's spatial domain (*spatial involvement*), (3) players' awareness
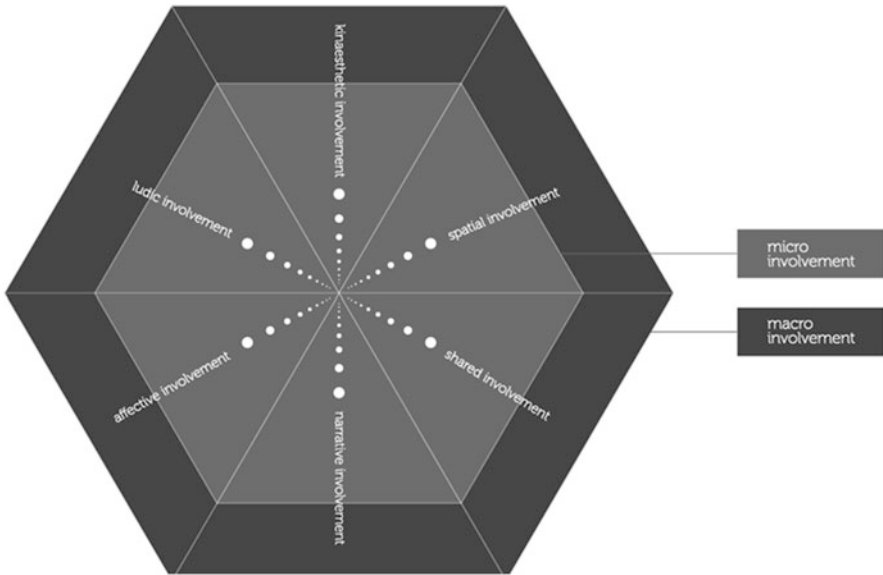
Fig. 3.2 The player involvement model

of and interaction with other agents and/or players in the game environment (*shared involvement*), (4) story elements that have been written into a game, or those that emerge from the player's interaction with the game (*narrative involvement*), (5) the various rules, goals and choices provided by a game (*ludic involvement*) and (6) the emotions that are generated during gameplay (*affective involvement*) [3].

As mentioned earlier, these dimensions can be considered relative to two temporal phases: the macro-involvement phase and the micro-involvement phase. *The macro phase of the model* deals with the off-line, long-term involvement of players with a game; their desire to engage with a specific game, the desire to return to it afterwards, and all off-line thinking that occurs in between game-playing sessions [3]. While the macro phase does constitute a crucial part of the player experience, we will limit ourselves here to analyzing the micro-involvement phase due to the limited scope of this chapter. *Micro-involvement*, then, describes players' moment-to-moment (and thus immediate) engagement while playing a game. This temporal phase thus deals with the imminent quality of involvement during game-play [3]. As we discussed earlier, a pre-requisite to experiencing micro-involvement is the direction of attention to the game.

These phases tend to be experienced in a combinatory manner and should thus be seen as layered and transparent in nature. This means that one phase influences how another is experienced and interacted with. The dimensions of the Player Involvement Model similarly combine in experience, with the inclusion or exclusion
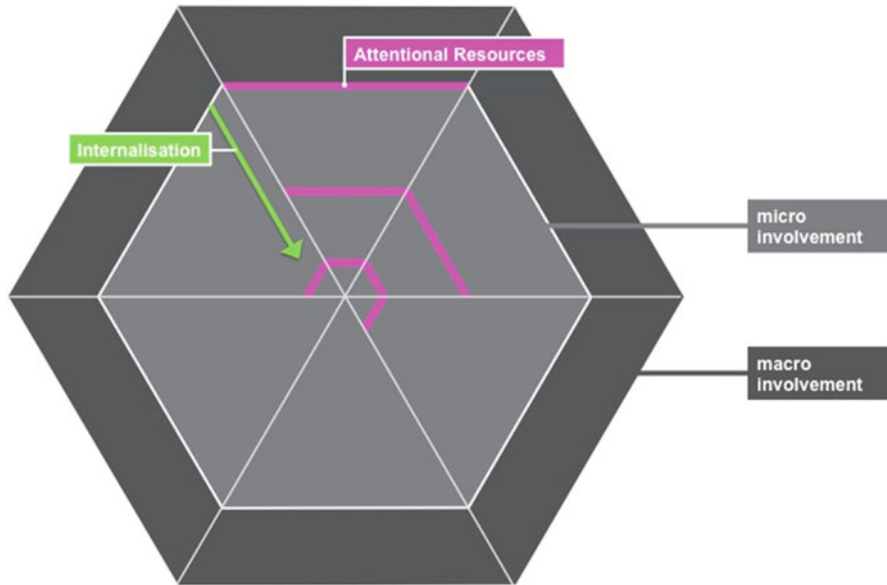
Fig. 3.3 Internalisation of involvement dimensions

of a dimension affecting how others are experienced [3]. The combinatorial aspect of the model makes it particularly useful for developing a structural approach to understanding game affect.

For instance, with Fig. 3.3 as a visual aid, we can think of the maximum attentional resources players have at their disposal as the outer line of each dimension's triangle. If players are dedicating all of their attention to a single dimension, then they will not be able to combine dimensions. If all of the attentional resources of the players are directed towards learning a game's controls, for example, the only dimension the players will interact with is kinesthetic involvement. As the controls are learnt, less attentional resources are demanded by kinesthetic involvement, allowing the players to pay attention to, for example, the actions of other members on their team (i.e. combining kinesthetic involvement with shared involvement), the story of the game (i.e. combining kinesthetic involvement with narrative involvement), etcetera.

The rest of the chapter will put into action this combinatorial nature of the Player Involvement Model [3], especially focusing on *combinations with affective involvement* in order to give a robust account of the forms of emotional types that games can elicit in our moment-to-moment engagement with them. Combining the affective involvement dimension with the other involvement dimensions will thus yield an experiential framework of game emotions that is unique to the gaming situation.

A Quantitative Perspective

As we mentioned earlier, Calleja's [3] Player Involvement Model was developed through extensive qualitative research. Building on this theoretical framework, a necessary subsequent step is to test the model in an experimental context, in order to validate, disprove or complement the model in a *quantitative set-up*. This will provide a new foundation for future research studying the player experience. Therefore, the aim of the current study was to test the micro phase of Calleja's [3] Player Involvement Model experimentally, investigating how the different dimensions of player involvement relate to, react and combine with each other. In light of the current chapter, we pay specific attention to affective involvement.

In order to be able to quantitatively examine the combinatorial nature of the Player Involvement Model, we conducted *three experiments*. The aim of these experiments was to test how varying one or two dimension(s) of player involvement – by manipulating a specific component of a digital game (e.g. game story, social setting, game control) – affects the other dimensions, and to what degree the different dimensions of player involvement are related to affective involvement.

Experimental Design

Experiment 1: Game Story

In our first experimental study, we focused primarily on *narrative involvement* (i.e. the player's involvement with the story elements that have been written into a game, and those that emerge from the player's interaction with the game) and *ludic involvement* (i.e. the dimension of player involvement concerned with the various rules, decision-making processes, goals and rewards of the game). These two dimensions are strongly connected: game stories do not only provide the player with background information, they outline the context of the game and often point out the player's objectives and the rewards he will get upon completing them. In turn, the goals of a game do not only help players find their position in the game, but also guide them in understanding the importance of their actions and choices within the narrative setting [3].

In order to vary these dimensions of player involvement, we manipulated the story of the game used in the experiment as a between-subjects factor, resulting in two experimental conditions. People in one experimental condition played a game level that had an elaborate story and in which the player had to perform an emotionally engaging task (i.e. save children from vicious raiders; *elaborate story* condition), while the people in the other experimental condition played a game level that had a minimal story and in which the player only had to perform a simple, unemotional task (i.e. collecting objects; *minimal story* condition).

We used the PC version of the action role-playing game *Fallout*: *New Vegas* (Bethesda Softworks, 2010) for this purpose. Using the game's editor, we created

an experimental level that suited the purpose of the study. Only the game's story was manipulated between the experimental conditions, meaning that apart from the implementation of the game's narrative, participants in the different conditions played the exact same game level.

Sixty-two experienced gamers (57 male, 5 female) between 18 and 37 years old ($M = 22.32$, $SD = 3.21$) eventually participated in the experiment (i.e. 31 people per experimental condition). During the experiment, participants first played a tutorial level explaining the basics of the game, after which the actual experimental level would start. The experimental level took approximately 12 min to finish.

Experiment 2: Social Setting

In our second experiment, we addressed *shared involvement* (i.e. player involvement related to the awareness of and interaction with other agents or players in a game environment).

In order to vary this dimension of player involvement, we manipulated the social context in which participants played a digital game as a between-subjects factor, leading to four experimental conditions. People either played the experimental game alone (*single-player* condition); together with one other person (*multiplayer* condition); or alone while another person watched (*public play* condition). Because this last condition consists of both players and observers, it is further divided into two groups (*public play*: *player* and *public play*: *observer* groups).

For this purpose, the PlayStation 3 puzzle-platformer game *LittleBigPlanet 2* (Sony Computer Entertainment, 2010) was used. By using the editor of the game, we created our own experimental level. Only the social setting in which the game was played differed between conditions, meaning all participants played the exact same level.

One hundred twenty-one gamers (82 male, 39 female), 18–24 years of age ($M = 20.69$, $SD = 1.79$) participated in the experiment. The single-player condition included 31 participants, while the three other conditions all contained 30 participants. During the experiment, participants first played a tutorial level that explained the basics of the game, after which the experimental level started. The experimental level had an average play time of 8 min.

Experiment 3: Game Controller

For the third experimental study we looked at *kinesthetic involvement* (i.e. player involvement related to control and movement in the game environment) and *spatial involvement* (i.e. player involvement with the exploration, navigation and learning of the game's spatial domain). Again, these two dimensions are strongly intertwined: players have to control a game in order for their character(s) on-screen to be capable

of movement and actually navigate through the game space. Until players learn to move in the world, they cannot engage with its spatial dimensions [3].

In order to vary these dimensions, we performed a within-subjects experiment in which we manipulated the game controller that was used in two experimental conditions. This means that participants played the same game twice, once with a PlayStation 3 gamepad controller (i.e. traditional controller), and once with a PlayStation Move racing wheel controller (i.e. motion-based controller).

The PlayStation 3 kart-racing game *LittleBigPlanet Karting* (Sony Computer Entertainment, 2012) was employed in the experiment. By using the editor of the game, we created an experimental level for use in the study. Only the game controller that was used differed between conditions, meaning participants played the exact same game level twice, each time with a different controller.

Thirty-one gamers (24 male, 7 female), 18–30 years of age ($M = 22.61$, $SD = 2.99$) participated in the within-subjects experiment. During the experiment, participants first played the official tutorial level of the game, after which the experimental level started. This experimental level had an average play time of 6 min.

Measures

After participants finished playing the games in the experiments, they were asked to fill out a questionnaire asking them how involved (on all dimensions) they felt during gameplay.

Narrative involvement was measured by a combination of narrative engagement scales from Busselle and Bilandzic [1], de Graaf et al. [5] and Green and Brock [8], adapted for use in a game context where necessary. The scale includes statements regarding the player's *focus on and interest in the story* (e.g. "My attention was fully captured by the story", "I was interested in the game's story"; 6 items, *Cronbach's α* = 0.85). Agreement with these statements was measured on 5-point intensity scales ranging from 0 ("not at all") to 4 ("extremely").

The *ludic involvement* scale was based on a combination of items from IJsselsteijn et al.' [9] Game Experience Questionnaire and de Graaf et al.'s [5] narrative engagement scale. It includes statements assessing the player's *interest in and focus on the game's mission and goals* (e.g. "I was fully concentrated on reaching the goals of the mission", "I was interested in the mission of the game"; five items, *Cronbach's α* = 0.68). Agreement was measured on 5-point intensity scales ranging from 0 ("not at all") to 4 ("extremely").

Shared involvement was measured by making use of the Social Presence module of IJsselsteijn et al.'s [9] Game Experience Questionnaire. The scale includes statements regarding *empathy with the other* and *behavioral involvement* (e.g. "I felt connected to the other", "My actions depended on the other's actions"; 14 items, *Cronbach's α* = 0.83) to which agreement is measured on a 5-point intensity scale ranging from 0 ("not at all") to 4 ("extremely").

Kinesthetic involvement was measured by making use of a combination of items from previously validated (kinesthetic) involvement questionnaires such as Witmer and Singer's [17] Presence Questionnaire, Jennett et al.'s [10] Immersion Scale and Vanden Abeele's [16] Perceived Control Scales. The utilized scale includes statements regarding the player's perceived amount of *control over the actions and movements of the avatar in-game* and the *difficulty of the game controls* (e.g. "I felt that I was able to control what happened in the game environment", "The game controls were easy to learn"; nine items, *Cronbach's* $\alpha = 0.94$). Agreement with these items was measured on a 5-point intensity scale ranging from 0 ("not at all") to 4 ("extremely").

Spatial involvement was measured by making use of items from IJsselsteijn et al.'s [9] Game Experience Questionnaire, in combination with items based on descriptions of the dimension given by Calleja [3]. The scale includes statements regarding the *exploration of the game level* and the *navigation through the game world* (e.g. "I felt that I could explore and discover things", "The navigation through the game level was difficult"; five items, *Cronbach's* $\alpha = 0.67$) to which agreement is measured on a 5-point intensity scale ranging from 0 ("not at all") to 4 ("extremely").

Affective involvement was measured by taking into account both general emotions (i.e. largely uncontrollable and spontaneous emotional reactions that are continuously present to some degree), and more specific player experiences.

When people play digital games, they experience several general emotions, such as pleasure, arousal and dominance. *Pleasure* refers to the pleasantness or enjoyment of a certain experience [14], *arousal* gives an indication of the level of physical and mental activation associated it [14], and *dominance* concerns the feeling of control and influence over others and surroundings [11]. The general emotions participants felt while playing the game were measured by using Lang's [12] Self-Assessment Manikin. This scale uses 9-point visual scales ranging from 0 to 8, with ascending scores corresponding to higher pleasure, arousal and dominance ratings.

Apart from general emotions, digital games have the potential to evoke a wealth of *specific player experiences*, such as challenge, competence, tension, positive affect and negative affect [10, 13]. These player experiences were also measured using IJsselsteijn et al.'s [9] Game Experience Questionnaire. The Game Experience Questionnaire includes statements to which agreement is measured on a 5-point intensity scale ranging from 0 ("not at all") to 4 ("extremely"). *Challenge* measures the stimulation players perceive and the amount of effort they have to put into the game (e.g. "I felt challenged", "I felt stimulated"; five items, *Cronbach's* $\alpha = 0.64$–0.82). *Competence* refers to how successful and skilful people feel while playing a game (e.g. "I felt successful", "I felt skilful"; five items, *Cronbach's* $\alpha = 0.87$–0.96). *Tension* measures the degree to which players feel frustrated and annoyed (e.g. "I felt frustrated", "I felt irritable"; three items, *Cronbach's* $\alpha = 0.66$–0.83). *Positive affect* probes players' fun and enjoyment of the game (e.g. "I felt good", "I enjoyed it"; five items, *Cronbach's* $\alpha = 0.79$–0.88). *Negative affect* is concerned with the degree to which players are feeling bored and distracted (e.g. "I felt bored", "I found it tiresome"; four items, *Cronbach's* $\alpha = 0.64$–0.74).

Results

Impact of Manipulations on Player Involvement

To be able to analyze the impact of several components of digital gameplay on the dimensions of player involvement, as well as the relationships that exist between these dimensions, we conducted three experiments.

In what follows, we will first examine how varying one or two dimension(s) of player involvement (i.e. by manipulating a game's story, social setting or game controller) influences the other dimensions of involvement as well, with a particular focus on affective involvement.

Experiment 1: Game Story

Our first experiment manipulated the story of the game participants had to play as a between-subjects factor, resulting in two experimental conditions: an *elaborate story condition* and a *minimal story condition*. In order to test the impact of this manipulation on player involvement in its narrative, ludic and affective dimensions, we performed one-way analyses of variance (ANOVAs).

Results of these analyses show that there are indeed significant differences in *narrative involvement* between our story conditions ($F(1, 60) = 5.13$, $p = 0.03$). Playing the game with the elaborate story led to significantly greater interest in and focus on the story ($M = 2.12$, $SD = 0.97$) compared to playing the game with the minimal story ($M = 1.63$, $SD = 0.71$).

However, contrary to our expectations, our results do not show significant differences in *ludic involvement* ($F(1, 60) = 0.04$, $p = 0.84$) or *affective involvement* (*pleasure*: $F(1, 60) = 2.59$, $p = 0.11$; *arousal*: $F(1, 60) = 0.02$, $p = 0.88$; *dominance*: $F(1, 60) = 0.12$, $p = 0.73$; *competence*: $F(1, 60) = 0.12$, $p = 0.73$; *tension*: $F(1, 60) = 0.14$, $p = 0.71$; *challenge*: $F(1, 60) = 0.04$, $p = 0.84$; *negative affect*: $F(1, 60) = 0.01$, $p = 0.93$; *positive affect*: $F(1, 60) = 2.10$, $p = 0.15$) between conditions. Players were not more or less absorbed in the ludic and affective dimensions of the game when playing with the elaborate versus minimal story.

We believe that, for this study, our manipulation of game story may not have been extensive enough to lead to differences in other dimensions between experimental conditions. An alternative explanation may be that some players are simply more focused on the narrative of the game than others. In a game, it is often up to the players themselves to decide whether they want to engage with the entirety of the story or simply focus on the goal-oriented tasks that push the game forward. It may be that in the current experimental setting, the narrative disposition of each of the players had a more important part to play in affecting player involvement than the between-subjects manipulation of the game story. In the section on *Combining affective involvement* (see further), we will take a look at potential relationships between narrative, ludic and affective involvement across conditions.

Experiment 2: Social Setting

The second experiment manipulated the social setting in which participants played a game as a between-subjects factor, resulting in four experimental conditions: a *single-player condition*, a *multiplayer condition*, a *public play*: *player condition* and a *public play*: *observer condition*). To be able to analyze the impact of this manipulation on player involvement, we again conducted one-way ANOVAs.

The results of these analyses demonstrate that there are significant variations in *shared involvement* between conditions ($F(2, 86) = 11.43$, $p < 0.001$), with people in the multiplayer group experiencing the highest sense of shared involvement ($M = 2.51$, $SD = 0.53$), followed by people in the public play: player ($M = 2.16$, $SD = 0.57$) and public play: observer groups ($M = 1.82$, $SD = 0.55$). This seems logical, since the multiplayer setting can be seen as the most social condition, with players having to actively interact with each other and synchronize their behavior in order to be able to work together. The public play: player setting can be considered as less social; although participants in this group played the game in a social setting (i.e. while being watched), the actual gameplay was experienced solo. However, this setting is still more susceptible to social influence compared to the public play: observer condition, since participants in the latter group did not have to actually perform the task (i.e. controlling the game).[1]

Moreover, our results demonstrate that the variations in social setting had a significant impact on *affective involvement*, especially affecting *arousal* ($F(3, 117) = 3.90$, $p = 0.01$), *dominance* ($F(3, 117) = 6.83$, $p < 0.001$) and *negative affect* ($F(3, 117) = 3.61$, $p = 0.02$). Participants in the multiplayer context felt more aroused ($M = 4.53$, $SD = 1.59$) than participants in public play: player ($M = 3.83$, $SD = 1.42$), public play: observer ($M = 3.73$, $SD = 1.46$) and single-player context ($M = 3.16$, $SD = 1.79$). However, people in the single-player condition reported higher levels of dominance ($M = 5.26$, $SD = 1.41$) than people in the public play: player ($M = 4.63$, $SD = 1.30$), multiplayer ($M = 4.27$, $SD = 1.39$) and public play: observer conditions ($M = 3.73$, $SD = 1.31$). Moreover, negative affect, although low in all conditions, was highest in the public play: observer condition ($M = 0.33$, $SD = 0.41$), differing significantly from the experienced negative affect in the other conditions (single-player: $M = 0.14$, $SD = 0.33$; multiplayer: $M = 0.09$, $SD = 0.28$; public play: player: $M = 0.12$, $SD = 0.22$).

A possible explanation for these results may be that players in the social play conditions (and especially the multiplayer condition) were more excited to be playing with/in the presence of another person, but did not have enough time to balance their individual playing and/or communication styles during the short period in which the experiment took place, resulting in a clumsy collaboration and the players feeling less dominant. Moreover, observers experienced less dominance

---

[1]Players in the single-player condition were not questioned about their shared involvement because they participated in the experiment alone. Therefore, they could not be aware of or interact with others.

than players, and more negative affect. This result can be attributed to the fact that observers are not in control of what happens in the game and therefore experience emotions that are less intense (i.e. dominance) on the one hand, and more boredom (i.e. negative affect) on the other.

Experiment 3: Game Controller

The third and final experiment manipulated the type of game controller that was used to play a game as a within-subjects factor, resulting in two experimental conditions: a *traditional game controller condition* and a *motion-based game controller condition*. In order to analyze the impact of this manipulation on player involvement, we performed repeated measures ANOVAs.

Our results show that manipulating the game controller has a significant impact on *kinesthetic involvement* ($F(1, 30) = 81.78$, $p < 0.001$): the controls of the traditional controller were perceived as easier to learn and handle ($M = 3.04$, $SD = 0.76$) than those of the motion-based controller ($M = 1.55$, $SD = 0.70$), allowing the players more precise control over their movements and actions in the game world.

Further, the findings also reveal a significant impact of the manipulation of game controller on *spatial involvement* ($F(1, 30) = 27.48$, $p < 0.001$); when playing the game with the traditional controller, the exploration and navigation of the game's spatial domain was perceived to be more easy and uncomplicated ($M = 2.87$, $SD = 0.51$) compared to playing with the motion-based controller ($M = 2.21$, $SD = 0.76$).

Finally, the game controller that was used to play the game was shown to significantly influence *affective involvement*. First of all, the manipulation of game controller significantly affected the players' sense of *dominance* ($F(1, 30) = 34.70$, $p < 0.001$): playing with the traditional game controller resulted in people feeling more dominant and in control ($M = 5.58$, $SD = 1.54$) compared to playing with the motion-based game controller ($M = 3.45$, $SD = 1.43$). Moreover, results show a significant effect on *competence* ($F(1, 30) = 71.02$, $p < 0.001$), *tension* ($F(1, 30) = 25.49$, $p < 0.001$), *challenge* ($F(1, 30) = 65.48$, $p < 0.001$) and *positive affect* ($F(1, 30) = 14.86$, $p = 0.001$). Playing with the traditional controller resulted in more competence ($M_{\text{Traditional}} = 2.90$, $SD = 0.82$; $M_{\text{Motion-based}} = 1.47$, $SD = 1.02$) and positive affect ($M_{\text{Traditional}} = 2.81$, $SD = 0.60$; $M_{\text{Motion-based}} = 2.34$, $SD = 0.85$), and less tension ($M_{\text{Traditional}} = 0.53$, $SD = 0.59$; $M_{\text{Motion-based}} = 1.33$, $SD = 1.03$) and challenge ($M_{\text{Traditional}} = 1.01$, $SD = 0.72$; $M_{\text{Motion-based}} = 2.17$, $SD = 0.75$).

Combining Affective Involvement

The previous section demonstrates that manipulating a specific component of digital gameplay (i.e. game story, social setting, game control) has a significant impact on player involvement, not only affecting the obviously linked dimensions (e.g. social

setting was expected to influence shared involvement), but in two out of three cases also reflecting on affective involvement.

In the current section, we look at the relationships between affective involvement and the other dimensions of player involvement across conditions, in order to get a more detailed look at the interrelatedness of the six dimensions. The first experiment provides us with data regarding relations between affective, narrative and ludic involvement; the second experiment reveals relations between affective and shared involvement; and the third experiment exposes relations between affective, kinesthetic and spatial involvement.

Experiment 1: Game Story

In the first experiment, we found that manipulating the story of a game (i.e. elaborate story versus minimal story) led to significant differences in narrative involvement between conditions. Against our expectations, however, ludic and affective involvement were not significantly influenced. By making use of correlation analyses, we now take into account relationships between the respective dimensions of involvement across conditions.

Narrative Involvement and Affective Involvement

The results of these correlation analyses reveal significant relationships between narrative involvement and several components of affective involvement. First, we observe moderate to strong positive relationships between narrative involvement and *pleasure* (Pearson's $r = 0.32$, $p = 0.01$), *arousal* (Pearson's $r = 0.31$, $p = 0.02$), *competence* (Pearson's $r = 0.43$, $p < 0.001$), *positive affect* (Pearson's $r = 0.45$, $p < 0.001$) and *challenge* (Pearson's $r = 0.47$, $p < 0.001$). Moreover, narrative involvement is shown to be moderately and negatively related to *negative affect* (Pearson's $r = -0.33$, $p = 0.01$).

Ludic Involvement and Affective Involvement

Further, correlation analyses also show significant relationships between ludic involvement and affective involvement, with ludic involvement being moderately and positively related to feelings of *competence* (Pearson's $r = 0.38$, $p = 0.003$) and *positive affect* (Pearson's $r = 0.30$, $p = 0.02$).

Experiment 2: Social Setting

The results of the second experiment showed us that varying the social setting in which a digital game is played (i.e. single-player setting, multiplayer setting, public

play: player setting, public play: observer setting) can significantly affect shared involvement, as well as lead to differences in player affect. Correlation analyses further provide us with a greater insight into the relationships between the two dimensions.

Shared Involvement and Affective Involvement

The results of correlation analyses demonstrate that shared involvement is significantly related to various aspects of affective involvement. More specifically, we record moderate to strong positive relationships between shared involvement and *pleasure* (Pearson's $r = 0.38$, $p < 0.001$), *arousal* (Pearson's $r = 0.30$, $p = 0.01$) and *positive affect* (Pearson's $r = 0.48$, $p < 0.001$), while moderate negative relationships are observed between shared involvement and *dominance* (Pearson's $r = -0.23$, $p = 0.03$) and *negative affect* (Pearson's $r = -0.33$, $p = 0.002$).

Experiment 3: Game Controller

Finally, the results of the third experiment demonstrated that manipulating the type of game controller that is used to play a game (i.e. the traditional PlayStation 3 gamepad controller versus the motion-based PlayStation Move racing wheel controller) has a significant impact on kinesthetic, spatial and affective involvement. Correlation analyses give us a more detailed look concerning the relationships between these dimensions.

Kinesthetic Involvement and Affective Involvement

The findings of these correlation analyses show that there are significant relationships between kinesthetic involvement and a variety of components of affective involvement. Kinesthetic involvement seems to be strongly and positively associated with *pleasure* (Pearson's $r = 0.52$, $p = 0.003$), *dominance* (Pearson's $r = 0.54$, $p = 0.002$), *competence* (Pearson's $r = 0.78$, $p < 0.001$) and *positive affect* (Pearson's $r = 0.46$, $p = 0.01$). Moreover, strong negative relationships are registered regarding kinesthetic involvement and *arousal* (Pearson's $r = -0.62$, $p < 0.001$), *tension* (Pearson's $r = -0.45$, $p = 0.01$) and *challenge* (Pearson's $r = -0.57$, $p = 0.001$).

Spatial Involvement and Affective Involvement

Finally, the correlation analyses reveal significant relationships between spatial involvement and affective involvement. Spatial involvement is strongly and pos-

itively related to *pleasure* (Pearson's $r = 0.66$, $p < 0.001$), *competence* (Pearson's $r = 0.52$, $p = 0.003$) and *positive affect* (Pearson's $r = 0.61$, $p < 0.001$); and strongly and negatively related to *negative affect* (Pearson's $r = -0.45$, $p = 0.01$) and *tension* (Pearson's $r = -0.49$, $p = 0.01$).

Conclusion

The current chapter takes as its main object of study the Player Involvement Model, an analytical framework designed to understand player experience developed through qualitative research by Calleja [3]. The Player Involvement Model identifies six dimensions of involvement in digital games, namely kinesthetic, spatial, shared, narrative, ludic and affective involvement. As the current collection focuses on emotions in games, it was our goal to elaborately investigate *affective involvement* and its relationship with the other dimensions of player involvement in a *quantitative context*.

In order to do this, *three experimental studies* were set up. Each of these experiments aimed to vary one or two dimension(s) of player involvement by manipulating a specific component of a digital game participants had to play. In the first experiment, we manipulated the *story of a game* in order to study narrative and ludic involvement. The second experiment manipulated the *social setting* in which a game was played to influence shared involvement. Finally, the third experiment manipulated the *game controller* that was used to play a game in order to affect kinesthetic and spatial involvement.

The results of these experiments show that the manipulations have a significant impact on player involvement, not only affecting the intended dimensions, but in two out of three cases also the emotions that participants experienced during gameplay (i.e. affective involvement). Subsequently, we further investigated the combinations of the dimensions of player involvement with affective involvement across conditions. Our findings suggest that players' affective involvement can be influenced by their interaction with each of the other dimensions of involvement.

In the context of the first experiment, several components of affective involvement are significantly related to both narrative and ludic involvement.

When players were more focused on and involved in the story of the game (i.e. *narrative involvement*), they were more aroused, they perceived the game to be more challenging and felt more competent. Moreover, they experienced more positively-valenced emotions (i.e. pleasure and enjoyment) and less negatively-valenced emotions (i.e. boredom). These results support previous findings by Calleja [4], who similarly describes the relationship between narrative and affective involvement. It is worth noting that this relationship works both ways: affective involvement makes narrative moments more memorable and significant, whether these moments arise from the story pre-scripted by the game's designers or generated through interaction with the game [3, 4]; while simultaneously engaging narrative creates a context of meaning that allows for more positive emotions to be experienced.

Furthermore, greater involvement with the goals and rewards of the game (i.e. *ludic involvement*) was associated with players feeling more competent and skillful during gameplay, as well as experiencing more positively-valenced emotions (i.e. enjoyment). Again, this is in line with Calleja's [3] qualitative research where participants commented on the sense of satisfaction they experienced from attaining game goals and reaping the related rewards.

The results of the second experiment demonstrate that shared involvement is also related to affective involvement. When players experienced more involvement because of a higher awareness of, and more interactions with other players (i.e. *shared involvement*) in the study, they were more aroused and experienced more positively-valenced emotions (i.e. pleasure and enjoyment) on the one hand, while also feeling less dominant and experiencing less negatively-valenced emotions (i.e. boredom) on the other. The presence of these negative emotions associated with shared involvement is also highlighted by Calleja's [3] research. Here, participants reported how collaboration with others can be intensely exciting and satisfying, if the members in the relevant team(s) manage to work together well, or very frustrating if things go wrong. For teams of players to collaborate properly they would have needed to play together for a considerable amount of time. In our experiment, players only had a short time to play the game and synchronize their individual playing and communication styles; it is thus understandable that collaboration in our experiment led to both negative and positive emotions related to shared involvement.

Finally, the third experiment shows that the dimensions of kinesthetic and spatial involvement are significantly associated with affective involvement as well.

When players experienced more involvement because of a higher sense of control over the actions and movements of their game character (i.e. *kinesthetic involvement*), they perceived themselves to be more dominant and in control, more competent and skillful in general, and they experienced more positively-valenced emotions (i.e. pleasure and enjoyment). Further, they also perceived the game to be less challenging, and experienced less arousal and negatively-valenced emotions (i.e. frustration). Participants from Calleja's [3] studies previously linked these positively-valenced emotions in reaction to mastering and internalizing the game controls with a strengthening of the link between player and avatar, which yielded a stronger sense of virtual environment habitation or incorporation [3], an experience that was described in positive terms by his participants.

Lastly, we found that higher involvement with the exploration, navigation and learning of the game's spatial domain (i.e. *spatial involvement*) was associated with players feeling more competent, more positively-valenced emotions (i.e. pleasure and enjoyment), and less negatively-valenced emotions (i.e. frustration and boredom). This relationship between spatial and affective involvement is reminiscent of Calleja's [3] argument that the internalization of spatial involvement turns abstract virtual space into familiar place. Participants in Calleja's research commented on the positive emotions that stem from this transition:

I am a pretty visual person. I drive by landmarks not by road names, so, I can visualize it pretty vividly. Though some of the in-between spaces are hazy cause I fly from Undercity to Tarren Mill now. It doesn't matter if it's the real world or digital, as I travel around, and learn new areas, I naturally seek certain kinds of landmarks to help me keep my bearings. A twist in the road here, a tree there. I find it comforting when I start to get the lay of the land in an MMO. For others, I have no idea. It's not just comfort, but also a bit of pride. I know where I am, I know where I want to go, and I know how to get there. (Rheric, quoted in [3])

Based on these findings, we can provide initial support for Calleja's [3] Player Involvement Model in a quantitative setting. Our experiments show the existence of the six dimensions of player involvement, and demonstrate that a manipulation of certain components of digital gameplay can result in significant effects on these dimensions. Moreover, our results reveal that the dimensions of player involvement are indeed interrelated, and that affective involvement can react to and combine with each of the other dimensions in different gaming situations.

Nevertheless, further research is still essential in order to come to a better understanding of players' involvement in digital games. Our studies have examined involvement in games such as action role-playing game *Fallout: New Vegas* (Bethesda Softworks, 2010), puzzle-platformer *LittleBigPlanet 2* (Sony Computer Entertainment, 2010) and kart-racing game *LittleBigPlanet Karting* (Sony Computer Entertainment, 2012), in a variety of digital game situations. However, a lot of different game genres, platforms, controllers, social modes, . . . exist, all possibly affecting the player in different ways, both cognitively, affectively and conatively. It is therefore of high importance that future studies continue to investigate player involvement in a multitude of digital games and gaming situations.

Finally, it is important to note that there is not yet a reliable, multidimensional scale of player involvement available for use in quantitative research. The experiments described in this chapter made use of a combination of previously validated (player) experience scales, adapted for use in the current studies. Although this suited the purpose of our initial and explorative studies, the development of a reliable, valid, sensitive and multidimensional measure based on Calleja's [3] Player Involvement Model would be very valuable for game researchers studying player experience. In that regard, following Yannakakis and Martinez [18], a rank-based questionnaire instead of a rating-based questionnaire to tap into player involvement could be considered. This would overcome some serious shortcomings (e.g. ordinal values treated as numerical, non-linearity of the scale) inherent to the use and analysis of the self-report items in the current studies.

References

1. Busselle R, Bilandzic H (2009) Measuring narrative engagement. Media Psychol 12(4): 321–347
2. Brown E, Cairns P (2004) A grounded investigation of immersion in games. CHI 2004, Vienna, Austria

3. Calleja G (2011) In-game: from immersion to incorporation. MIT Press, Cambridge, MA
4. Calleja G (2013) Experiential narrative. Proceedings of the International Conference for the Foundation of Digital Games 2013. Chania, Greece
5. de Graaf A, Hoeken H, Sanders J, Beentjes H (2009) The role of dimensions of narrative engagement in narrative persuasion. Communications 34(2009):385–405
6. Douglas YJ, Hargadon A (2001) The pleasure of immersion and engagement: schemas, scripts and the fifth business. Digit Creat 12(3):153–166
7. Dovey J, Kennedy HW (2006) Game cultures: computer games as new media. Open University Press, Berkshire
8. Green MC, Brock TC (2000) The role of transportation in the persuasiveness of public narratives. J Pers Soc Psychol 79(5):701–721
9. IJsselsteijn WA, de Kort YAW, Poels K (2008) The game experience questionnaire: development of a self-report measure to assess player experiences of digital games. FUGA Deliverable D3.3. Eindhoven University of Technology.
10. Jennett C, Cox AL, Cairns P, Dhoparee S, Epps A, Tijs T, Walton A (2008) Measuring and defining the experience of immersion in games. Int J Hum Comput Stud 66(2008):641–661
11. Klimmt C, Hartmann T, Frey A (2007) Effectance and control as determinants of video game enjoyment. CyberPsychol Behav 10(6):845–847
12. Lang PJ (1980) Behavioral treatment and bio-behavioral assessment: computer applications. In: Sidowski JB, Johnson JH, Williams TA (eds) Technology in mental health care delivery systems. Alex, New Jersey, pp 119–137
13. Poels K, de Kort YAW, IJsselsteijn WA (2012) Identification and categorization of digital game experiences: a qualitative study integrating theoretical insights and player perspectives. Westminster Papers in Communication and Culture, 9(1), Special Issue: Encountering the Real Virtuality: Digital Games in Media, Culture and Society
14. Ravaja N, Saari T, Laarni J, Kallinen K, Salminen M (2005) The psychophysiology of video gaming: phasic emotional responses to game events. Retrieved from: http://www.digra.org/dl/db/06278.36196.pdf
15. Salen K, Zimmerman E (2003) Rules of play: game design fundamentals. MIT Press, Cambridge, MA
16. Vanden Abeele V (2011) Motives for motion-based play. Less flow, more fun. Doctoral dissertation. Katholieke Universiteit Leuven
17. Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: a presence questionnaire. Presence 7(3):225–240
18. Yannakakis GN, Martinez HP (2015) Ratings are overrated! Front ICT 2(13)

Part II
Emotion Modelling and Affect-Driven Adaptation

Chapter 4
Multimodal Sensing in Affective Gaming

Irene Kotsia, Stefanos Zafeiriou, George Goudelis, Ioannis Patras, and Kostas Karpouzis

Abstract A typical gaming scenario, as developed in the past 20 years, involves a player interacting with a game using a specialized input device, such as a joystick, a mouse, a keyboard or a proprietary game controller. Recent technological advances have enabled the introduction of more elaborated approaches in which the player is able to interact with the game using body pose, facial expressions, actions, even physiological signals. The future lies in 'affective gaming', that is games that will be 'intelligent' enough not only to extract the player's commands provided by speech and gestures, but also to extract behavioural cues, as well as emotional states and adjust the game narrative accordingly, in order to ensure more realistic and satisfactory player experience. In this chapter, we review the area of affective gaming by describing existing approaches and discussing recent technological advances. More precisely, we first elaborate on different sources of affect information in games and proceed with issues such as the affective evaluation of players and affective interaction in games. We summarize the existing commercial affective gaming applications and introduce new gaming scenarios.

I. Kotsia (✉)
Department of Computer Science, Middlesex University, London, UK
e-mail: irene.kotsia@eecs.qmul.ac.uk;

S. Zafeiriou
Department of Computing, Imperial College London, London, UK
e-mail: s.zafeiriou@imperial.ac.uk

G. Goudelis
Image, Video and Multimedia Systems Lab, National Technical University of Athens,
Athens, Greece
e-mail: ggoudelis@image.ntua.gr

I. Patras
School of Electronic Engineering and Computer Science, Queen Mary University of London,
London, UK
e-mail: i.patras@eecs.qmul.ac.uk

K. Karpouzis
Institute of Communication and Computer Systems, National Technical University of Athens,
Zographou, Greece
e-mail: kkarpou@image.ntua.gr

We outline some of the most important problems that have to be tackled in order to create more realistic and efficient interactions between players and games and conclude by highlighting the challenges such systems must overcome.
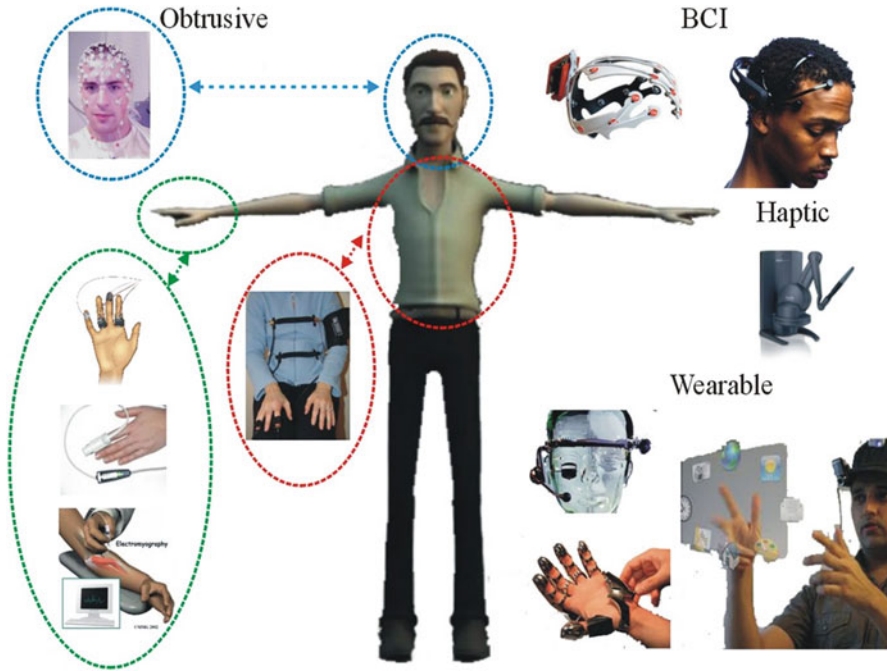
Introduction

The games industry has grown to be one of the mainstream markets in our days. In the beginning, the games industry constituted a focused market, highly depended on specialized input sensors to enable the interaction between a player and the game. In typical games, the user had to be familiar with an input device, such as a keyboard, a mouse, a joystic or a console, in order to properly communicate with the game. Furthermore, the game had a predesigned plot that would progress along with the players actions in a predefined way, giving the feeling of a non-existent, in reality, control over how the game evolves. Moreover, in such a gaming scenario several issues had to be tackled: the game had to be carefully designed and developed so as to allow real-time interaction with the player, ensure a high quality visual environment, so that the immersion of the player in the game environment would be as realistic as possible, and employ devices that would be of affordable cost.

Initial research approaches in the field of affective computing focused on processing the physiological cues of a player in order to correlate them with certain behavioural patterns that would assist in making the player-game interaction more realistic and meaningful. To achieve that, several physiological signals were employed, such as heart beat rate, skin conductivity etc., using obtrusive devices. The use of brain signals also defined a field on its own, leading in the creation of Brain Computer Interface (BCI) systems. The most recent approaches tried to create wearable systems that were built of portable devices/personal computers and thus eliminated the effect of sensors as they provided the player with extra degrees of freedom. However, the main problem with employing specialized sensors to extract behavioural cues is that they greatly affect the immersion of the player in the game. Even with sensors that are relatively easy to use, such as skin conductivity sensors, the player's actions are constrained by the space limitations of each sensor. This is of great importance as it usually leads the player to exhibit unusual behavioural patterns often attributed to the effect (even subconscious one) that the presence of a sensor has. An overview of the available sources of information is depicted in Fig. 4.1, along with the corresponding part of the body from which physiological signals are extracted.

Recent technological advances have opened new avenues towards more realistic human-machine interaction systems, thus enabling the introduction of games in which the player is not only able to control the game, but can also control the gameplot and player experience without using an input device, just by freely acting on his/her will. More specifically, the introduction of Microsoft Kinect [57] has enabled the robust extraction of body poses and joint locations in real-time, while at the same time being of low cost. By providing such a real time and robust solution to the tracking problem, research has been shifted towards affective gaming scenarios,

Fig. 4.1 An overview of the widely used sensors

in which the player's emotional states and actions will be used to define the way the game will progress.

The term affective gaming corresponds to the introduction of affect recognition in games. These approaches tried to incorporate emotion recognition in their systems in order to extract the emotional state of the player and use it to control the gameplot and gameplay experience. In this paper we will review the area of affective gaming. More precisely, we will briefly review existing approaches and discuss the recent technological advances. We will present the existing means for the extraction of behavioural cues and subsequently analyze existing approaches in affective gaming. Finally we will examine future trends in games by defining new affective gaming scenarios and also discussing their possible applications.

The remainder of the chapter is organized as follows. We first discuss the term affective gaming and present in brief its involvement through time (Section "Affective Gaming"). We continue with presenting the different sources of affect information, categorizing them in those that involve vision based techniques (Section "Vision-Based"), those that involve haptics as an input and interaction modality (Section "Haptics") and those that employ specialized wearable devices (Section "Wearable Games"). We proceed with investigating several issues raised in affective gaming, such as affective evaluation of players (Section "Affective Evaluation of Players") and affective interaction in games (Section "Affective Interaction in Games"). We also summarize the existing affecting gaming commercial applications in section "Applications of Affective Games". We introduce new affective gaming

scenarios (Section "Affective Gaming Scenarios") and discuss the challenges that affecting gaming systems must overcome (Section "Affective Gaming Challenges"). In section "Conclusions" we draw our conclusions.

Affective Gaming

The term affective gaming refers to the new generation of games in which the players' behaviour directly affects the game objectives and gameplay. More precisely, the emotional state and actions of a player can be correctly recognized and properly used in order to alter the gameplot and offer to the player an increased user experience feeling. In other words, the emotions and actions of a player are of extreme importance, as the behavioural cues extracted from them will define the way the game will progress.

Early approaches in the field of affective gaming focused on highlighting the importance of including emotional content in systems in order to make the user experience more satisfactory. Here the term 'emotion' corresponds to a variety of affective factors. Emotions are defined as short states (often lasting from seconds to minutes) that reflect a particular affective assessment of the state or self or the world and are associated with behavioural tendencies and cognitive biases [39]. They are further distinguished in universal (anger, disgust, fear, happiness, sadness and surprise [26]) and complex (guilt, pride and shame). Emotions are often defined in terms of their roles, thus being distinguished in those involved in interpersonal, social behaviour (e.g., communication of intent, coordination of group behaviour, attachment), and those involved in intrapsychic regulation, adaptive behaviour, and motivation (e.g., homeostasis, goal management, coordination of multiple systems necessary for action, fast selection of appropriate adaptive behaviours). They are usually manifested across four interacting modalities: the—most visible— behavioural/expressive modality (e.g., facial expressions, speech, gestures, posture, and behavioural choices), the somatic/physiological modality—the neurophysio- logical substrate making behaviour (and cognition) possible (e.g., changes in the neuroendocrine systems and their manifestations, such as blood pressure and heart rate), the cognitive/interpretive modality, directly associated with the evaluation- based definition of emotions and the experiential/subjective modality: the conscious, and inherently idiosyncratic, experience of emotions within the individual [39]. Emotions are most commonly characterized by two dimensions: valence and arousal [69]. The dimension of valence ranges from highly positive to highly negative, whereas the dimension of arousal ranges from calming or soothing to exciting or agitating. Affective computing has been directly linked with affective gaming in terms of emotion sensing and recognition, computational models of emotion, and emotion expression by synthetic agents and robots.

Emotion affect recognition has been widely researched in order to create human computer interaction systems that can sense, recognize and respond to the human communication of emotion, especially affective states such as frustration, confusion, interest, distress, anger, and joy. In [67], Picard highlighted the importance of recog-

nizing affective states commonly expressed around computer systems: frustration, confusion, dislike, like, interest, boredom, fear, distress, and joy. If a system is able to successfully detect expressions of these states, and associate them with its functions and other environmental events, then it can proceed with improving its interactions with the user. In that way the gap between the human and the machine in HCI systems can be narrowed and more user-centered HCI systems can be created [38]. Affective arousal modulates all nonverbal communicative cues (facial expressions, body movements, and vocal and physiological reactions), making efficient affect recognition a core element for emotionally intelligent systems [23, 63]. Interested readers can refer to [93] and [18] for a survey on affect recognition methods proposed through the years. In [32] the authors proposed a system for affect recognition that combines face and body cues in order to achieve affect recognition. They employed computer vision techniques (Hidden Markov Models (HMMs), Support Vector Machines (SVMs) and Adaboost) to fuse the available facial (appearance, e.g. wrinkles or geometric feature points) and body (silhouette and color based model) cues. However, the database on which the experiments were conducted included recordings of subjects sitting in front of a camera and reacting to a provided scenario. Therefore, although the subjects were free to express themselves, no real body pose information was available. Moreover, the database included recordings of a single person, thus not being suitable for group behaviour research.

Early approaches also focused on defining the fundamentals of affective gaming from a physiological point of view. The link between neurobiological perspectives and models of play aiming to construct superior player satisfaction models built upon biological foundations has been presented in [7]. More precisely, it was found that connections exist between already recognized patterns of play and recent research on the brain (in particular, the limbic system). In [31], Gilleade et al. discuss some of the origins of the genre, how affective videogames operate, and their current conceptual and technological capabilities. Early biofeedback-based affective games were regarded and a novel approach to game design based on several high-level design heuristics was proposed. In [40] the author discussed the enhancement of social and affective complexity and realism of the game characters of a game, their interaction and the game narrative as a whole. A set of requirements for an affective game engine, capable of supporting the development of more affectively realistic, engaging, and effective games was also proposed.

In [42] the authors proposed an Affect and Belief Adaptive Interface System designed to compensate for performance biases caused by users affective states and active beliefs. It implemented an adaptive methodology consisting of four steps: sensing/inferring user affective state and performance-relevant beliefs, identifying their potential impact on performance, selecting a compensatory strategy, and implementing this strategy in terms of specific GUI adaptations. In [24] the authors developed a computational framework for exploring cognitive decision processes that reason about emotional actions as an integral component of strategy and control over emotions. They implemented a prototype gaming system that exhibited conscious use of the emotional state and negative emotional behaviours during a game of chess in an attempt to impair its opponents game play. In [50], an educational chess game was developed with which the authors studied the role of emotions

and expressive behaviour in socially interactive characters employed in educational games. To this end they created a social robot named iCat functioning within a chess game scenario. In [39], Hudlicka discussed how affect recognition contributes to user experience and recognition of player emotion, and based on that, to tailoring game responses to recognised emotions, and generating affective behaviours in player and non-player characters. In [41] the link of affect recognition, as an efficient mean of generating appropriate behaviours in more complex environments, with cognition (attention, memory and language abilities) was also explored. Emotional states, moods and various other subjective experiences occurring just before a player engages with the game of or as a result of the gameplay or immediately after playing can be used to evaluate the player's experience [70] and possibly predict subsequent actions/decisions. Similar observations are made by Yannakakis and Togelius [91], where game content is generated procedurally, aimed to advance the player experience, while from the analysis point of view, Shaker et al. [75] investigate how affective expressivity, associated with game behaviour, can be used to predict player experience.

The ultimate goal of affective gaming systems is to create games that will be "intelligent" enough to understand what each player feels at each specific moment, using behavioural cues obtained either in an obtrusive way (e.g. using neurophysiological measurements), or in an unobtrusive one (e.g. observing facial expressions, body pose, actions, and behaviour). However, most existing scenarios employ sensors that limit the freedom of movement, for example the extraction of neurological cues requires the player to be sitting in front of the computer (game console). As a result, a poor quality/less realistic immersion of the player in the game environment may be achieved, due to the fact that the player may be conscious of the recording/measurement device and thus exhibit unusual behavioural patterns (not corresponding to spontaneous behaviour). Players tend to prefer less intrusive methods of physiological input for their games [60]. Therefore, the extraction of the necessary behavioural cues in an affective game scenario should be performed in a way that is not perceptible to players and does not limit their actions.

Affective gaming constitutes a field that aims at providing answers to a variety of different research questions. Issues regarding the player experience and how his/her affective and physiological evaluation can be performed, have to be discussed. Moreover, the measurement of interaction among players as well as the efficient modeling of their satisfaction remain open problems. Below we will briefly discuss all this issues in detail.

Sources of Affect

We will first describe the available sources for extracting affect information. These can be distinguished in three main categories: those that extract vision based information (such as facial expressions and body movements), those that extract information from brain signals (BCI applications) and those that extract physiological measurements employing specialized sensors.

Vision-Based

In this Section we will review the existing methods that extract affect information using vision-based techniques. This includes unobtrusive methods that do not limit the freedom of the player, allowing him/her to freely act on his/her will. These methods mainly attempt facial expression and body action recognition, as those are behavioural cues that can be easily obtained using a simple camera.

Facial Expressions

Although facial expressions recognition constitutes a well studied field on its own and its applications span several areas, among which games, a limited number of studies have been conducted targeted to affective gaming. An interested reader can refer to [71] and the references within for a brief introduction on recent advances of facial expressions and communicated affect.

In [6], the authors employ computer vision techniques to capture the player's behaviour in order to adapt the game and maximize the player enjoyment. More precisely, observed visual behaviour is associated with game events occuring at the same time; as a result, it is possible to reason about what caused each reaction (e.g. facial expression or facial cue) and whether that event was positive or negative with respect to the player's goal. This multimodal system makes it possible to differentiate between, for instance, an ironic smile when a player loses and a genuine smile after overcoming an opponent or finishing the level. In addition to face expressivity and head movement, body stance and movement also plays an important role in communicating a player's affective state. It has been shown that body expressions can reveal more information regarding the affective state of a person when nonverbal communication is considered [3].

In this section we look at some biological and cognitive aspects of facial expression recognition in humans. We should at this point stress that the subjective feeling of an emotion and its expression on the face are two different things, where the latter is one manifestation of the former among many bodily signals, like gestures, postures, and changes on the skin response. Thus, what we perceive from a face is either an involuntary manifestation of an emotional state, or the result of a deliberate effort at communicating an emotional signal. The urge to associate affect with faces is so great that we recognize expressions even on infants faces, even though they are not yet associated with the emotions they represent in adults. This association partly relies on innate biases implicit in the human visual system, and partly on the efficient way humans represent facial information. In humans, the subjective experience of an emotion, the production of its somatic expressions, and its recognition in other subjects are all tightly coupled, and influence each other. This allows for a degree of feedback that is beyond current computer systems, and enables differentiation of very subtle affective cues.

The goal of facial affect recognition systems is to mimic humans in their evaluations of facial expression. If a computer can learn to distinguish expressions automatically, it becomes possible to create interfaces that interpolate affective states from these expressions and use this information for better interfaces. We open a little parenthesis here. When we talk about learning in the context of a computer, we usually mean a machine learning procedure, which is different from human learning. Here, what usually happens is that the computer is provided with a number of samples from a category to be learned (be it images of faces with a particular expression or any other numeric representation), as well as a method of categorization. The learning algorithm tunes the parameters of the method to ensure a good categorization on these samples. The ensuing system, however, depends crucially on the quality of provided samples, in addition to the data representation, the generalization power of the learning method and its robustness to noise and incorrect labels in the provided samples. These points are shared by all computer systems working on face images, be it for the recognition of identity or expressions. We bear these in mind when investigating what the brain does with faces, and how it can be simulated with computers. Recognition of relevant processes that partake in human recognition of faces and facial affect guides the designers of computer algorithms for automatic recognition of emotions from faces. For instance, it is known that humans have selective attention for the eyes and mouth areas, which can be explained by recognizing the importance of these areas for communicating affect and identity. Computer simulations by [51] have shown that feature saliency for automatic algorithms that evaluate facial affect parallels feature saliency for the human visual system. How humans determine identity from faces is a widely researched area. One reason for this is that both low-level neurological studies and high-level behavioural studies point out to faces as having special status among other object recognition tasks. Kanwisher et al. [43] have argued that there is an innate mechanism to recognize faces, and they have isolated the lateral fusiform gyrus (also termed the fusiform face area) to be the seat of this process. The proponents of the expertise hypothesis, on the other hand, argued that humans process a lot of faces, and this is the sole reason that we end up with such a highly specialized system [30]. The expertise hypothesis banks on a fundamental property of the human brain: the key to learning is efficient representation, and while we learn to recognize faces, the neural representation of faces gradually changes, becoming tailored to the use of this information. In other words, we become (rather than born as) face experts. But this also means that we are sensitive to cultural particularities we are exposed to, an example of which is the famous other-race effect. This is also true for affect recognition from facial expressions, which incorporate cultural elements. While the geometric and structural properties of a face might allow the viewer to distinguish the basic emotional content, cross-cultural studies have established that the cultural background of the viewer plays a large role in labelling the emotion in a face. Furthermore, perception of emotion-specific information cued by facial images are also coloured by previous social experience. In a recent study [68], a number of children who have experienced a high-level of parental anger expression were shown sequences of facial expressions.

They were able to identify the anger expression in the sequence earlier than their peers, using a smaller amount of physiological cues. The traditional problems faced by face recognition researchers are illumination differences, pose differences, scale and resolution differences, and expressions. These variables change the appearance of the face, and make the task of comparing faces non-trivial for the computer. While there is a consensus among brain researchers that recognizing facial identity and facial affect involve different brain structures (e.g. lateral fusiform gyrus for identity as opposed to superior temporal sulcus for emotional content, [33]), these are not entirely independent [14]. Many aspects of facial identity recognition and affect recognition overlap. This is also the case for computer algorithms that are created for recognition of identity or affect from face images. Hence, it should be no surprise that computational studies also recognize the need for different, but overlapping representations for these two tasks. For instance Calder and colleagues [15] have investigated a popular projection based method for classifying facial expressions, and determined that the projection base selected to discriminate identity is very different than the base selected to discriminate expressions. Also, while facial identity concerns mostly static and structural properties of faces, dynamic aspects are found to be more relevant for emotion analysis. In particular, the exact timing of various parts of an emotional display is shown to be an important cue in distinguishing real and imitation expressions [22]. Similarly, the dichotomy of feature-based processing (i.e. processing selected facial areas) versus holistic processing (i.e. considering the face in its entirety) is of importance. Features seem to be more important for expressions, and while in some cases it can be shown that some expressions can be reliably determined by looking at a part of the face only [61], the dynamics of features and their relative coding (i.e. the holistic aspect) cannot be neglected. Before moving to tools and techniques for computer analysis of facial expressions, we note here that all emotions were not created equal. Brain studies suggest different coding mechanisms for particular emotions. According to the valence hypothesis there is a disparity between the processing of positive and negative emotions, as well as the amount of processing involved for these types in the left and right hemisphere of the brain [13]. This is an evolutionarily plausible scenario, as rapid motor response following particular emotions (e.g. fear, anger) is important for survival. Blair et al. [12] have found that the prefrontal cortex is more active for processing anger, as opposed to sadness. Different cortical structures show differential activation for different emotion types under lesion and functional imaging studies. On the other hand, specific emotions do share common neural circuitry, as disproportionate impairment in recognizing a particular emotion is very rare, as shown by lesion studies (the reader is referred to [1] for examples and references). This inequality is also reflected in displays of emotion. The configural distances from a neutral face are disproportionate for each emotion, with sadness and disgust being represented by more subtle changes (as opposed to for instance happiness and fear). In addition to this disparity, it is unlikely that emotions are encountered with the same background probability in everyday life. Thus, from a probabilistic point of view, it makes sense not to treat all six basic emotions on the same ground. The valence hypothesis suggests that happiness (as a positive

emotion) is a superordinate category, and should be pitted against negative emotions (fear, anger, disgust, sadness and contempt). Surprise can be divided into fearful surprise and pleasant surprise; it has been noted that surprise and fear are often confused in the absence of such distinction. Also, disgust encompasses responses to a large range of socially undesirable stimuli. When it expresses disapproval for other people, for instance, it approaches anger. These issues require careful attention in the design and evaluation of computer systems for facial expression analysis.

Body Expressivity

Changes in a persons affective state are also reflected by changes in body posture [56, 88], with some affective expressions being better communicated by the body than by the face [27, 28]. It has been shown that people tend to control facial expressions more than body expressions when trying to hide their emotions [28]. Conversely, people trust the expressions of the body more than the expressions of the face when the two are incongruent [34].

In [77, 78] the authors tried to categorize the affective behaviour of the child into a set of discrete categories by exploiting the information used by their body gestures. More precisely, they attached a motion capture system to the upper body part of the child in order to extract the coordinates of certain key points. A number of frames containing gesture information in the form of these points were selected and then used to an affective gesture recognition module that employed HMMs.

In [72] the authors extracted visual information of expressive postural features from videos capturing the behaviour of children playing chess with an iCat robot. To this end, they extracted the frontal and lateral view of the body of the subject (child) and used it to study several features, such as body lean angle, slouch factor, quantity of motion and contraction index. They classified these features using a set of different classifiers in order to identify the factors that reveal the player's engagement to the game.

In [19], Caridakis et al. utilised statistical measures of the movement of a user's hands and head to model affective body expression, and produced a five-dimensional representation termed "expressivity features". Despite the simplicity of this representation, which was robust to rotation and scaling or camera zooms, those dimensions (*overall activation*, *spatial extent*, *temporal*, *fluidity* and *power*) were found to be strongly correlated with human perception of affect in body expressivity in real (non-simulated) environments, showing that they can be utilised in human-computer interaction and game settings to perceive affective qualities from human movement, besides recognising actual gestures and postures. In [73], the authors used a similar approach, trying to recognize the affective states of players from non-acted, non-repeated body movements in the context of a video game scenario. To this end, they attached a motion capture system in order to collect the movements of the participants while playing a Nintendo Wii tennis game. They extracted several features from the body motion (*angular velocity*,

angular acceleration, *angular frequency*, *orientation*, *amount of movement*, *body directionality* and *angular rotations*) and the most discriminative of those were used as an input to a recurrent neural network algorithm. In the end they were able to recognize a set of eight emotions (frustration, anger, happiness, concentration, surprise, sadness, boredom and relief).

In [47, 74], the authors attempted recognition of affective states and affective dimensions from non-acted body postures instead of acted postures. The scenario they used was a body-movement-based video game (Nintendo Wii sports games). Postures were collected as 3-D joint Euler rotations of the joints and were used to recognize the affective state of the player after having won or lost a point. In these works, the following affective states were regarded: *concentrating* (determined, focused, interested); *defeated* (defeated, give up, sad), *frustrated* (angry, frustrated), and *triumphant* (confident, excited, motivated, happy, victory). The classification was performed using a multilayer perceptron network.

One of the well-known works is by Camurri et al. [16]. They examined cues involved in emotion expression in dance. The results ranged between 31 % and 46 % for automatically recognizing four emotions, whereas the recognition rate for the observers was considerably higher at 56 %. Berthouze et al. [48, 76] proposed a system that could recognize four basic emotions based on low-level features describing the distances between body joints. The system was tested on acted postures that were labeled by groups of observers from different cultures. The model built on observers that were from the same culture as most of the actors (Japanese) reached 90 % recognition. Using the same set of postures, similar performances were reached for observers from other cultures [48]. The same approach was used to automatically discriminate between affective dimensions, with performances ranging from 62 % to 97 %. Bernhardt and Robinson [8] built affect recognition models for nonstylized acted knocking motions using Pollick et al.'s database [52]. Their model takes into account individual idiosyncrasies to reduce the complexity of the modeling. After training, the classifier was tested on motion samples from a single actor. The results showed a 50 % recognition rate without removing the personal biases and 81 % with the biases removed. Picard's studies [44, 45] examine non-acted postures. Their multimodal system models a more complete description of the body, attempting to recognize discrete levels of interest [45] and self-reported frustration [44]. Of the input examined (facial expressions, body postures, and game state information), the highest recognition accuracy was obtained for posture (55.1 %). To conclude, most of the work has focused either on acted or stereotypical body expressions (e.g., dance), with the exception of the work presented in [45] and [44], where only a limited description of the body was considered. Low-level features have been shown to provide high-recognition performance. What is still lacking is an understanding of whether a similar approach can be used to create recognition models that automatically discriminate between natural expressions [88]. This is necessary to create automatic recognition systems that are usable in real contexts.

Haptics

Another mean widely used for feature acquisition in affective gaming scenarios is haptic technology. Haptics are devices that exploit the sense of touch by applying forces, vibrations, or motions to the user in order assist in the creation of virtual objects in a computer simulation, to control such virtual objects, and to enhance the remote control of machines and devices. The approaches that employed haptics in BCI systems are relatively new. More precisely, the first approach that employed haptics was presented in [62], in which the authors developed a low-cost and compact force feedback joystick as a new user interface to communicate with a computer. In order to achieve this, they adopted a joystick to create the illusion of force to human hand. They intercepted the interrupts reserved for mouse, colormap and keyboard input in a game and replaced them with a new interrupt service routine in which force effects are added for the new PC video games.

The use of haptics has been noticeably increased during the last decade. In the beginning it was used to aid people with disabilities. In [81] the authors presented a variety of simple games using haptics. The first one was a painting program for blind children, called 'Paint with your fingers'. The player used a haptic device to choose a color from a palette. Each color on the palette had an associated texture that the user could feel when painting with it. By changing program mode the user could feel the whole painting, and also print the painting on a colour printer. Another game, 'Submarines' was a haptic variant of the well known battleship game. The player had to feel 10×10 squares in a coordinate system. His/her finger in the haptic device was a helicopter that was hunting submarines with depth charge bombs. A third game, 'The Memory House' consisted of 25 pushbuttons that produced a sound when pressed. The buttons disappeared when the player pressed two buttons (using a haptic device) with the same sound in sequence. In the Memory House the buttons are placed on five different floors. Between each row of buttons the user can feel a thin wall that helps him to stay within one set of buttons. Much later, in [29] the authors discussed feedback in pervasive games and presented their work on Haptic Airkanoid, an extension to Airkanoid using Haptic. Airkanoid is a ball-and-paddle game where the player hit bricks in a wall with a ball. When a brick is hit, it disappeared and the ball was reflected. When all bricks were cleared the player was advanced to the next level. A user controlled paddle prevented the ball from getting lost. Graspable Airbats were used as interfaces for controlling the virtual paddles. In [65] the authors presented a Virtual Reality application of a game of billiard game that allowed the user to interactively provide a force feedback by means of a commercial haptic interface. The haptic device was used to simulate the skittle which the players used to hit the balls. In [66] the authors introduced a haptic interface for brick games. They used a haptic dial to add tactile feedback to enhance game effects in addition to visual and sound effects. The user was able to change the position of the paddle by spinning the dial knob while simultaneously feeling various tactile feedbacks according to the game context. In [17] the authors discussed a multipurpose system especially

suitable for blind and deafblind people playing chess or other board games over a network, therefore reducing their disability barrier. They used special interactive haptic device for online gaming providing a dual tactile feedback, thus ensuring not only a better game experience for everyone but also an improved quality of life for sight-impaired people. In [49] the authors created a multi-touch panel by using multi-touch technology in order to construct a suitable game interface and apply it to a game. In [90] the authors created a system model that helped dental students memorize fundamental knowledge as well as the processes and techniques in dental casting. To achieve that they have incorporated an Haptic interactive device in wireless vibration feedback for more lively and diverse learning methods in dental casting for learners. In [89] the authors measured playability of mobile games by comparing two different types of haptic interfaces, namely hard and soft keypad, for mobile gaming. In [37] the authors enhanced the open source Second Life viewer client in order to facilitate the communications of emotional feedbacks such as human touch, encouraging pat and comforting hug to the participating users through real-world haptic stimulation. In [79] the authors presented a haptic system for hand rehabilitation, that combined robotics and interactive virtual reality to facilitate repetitive performance of task specific exercises for patients recovering from neurological motor deficits. They also developed a virtual reality environment (maze game) in which the robot applied force fields to the user as the user navigated the environment, forming a haptic interface between the patient and the game. In [85] the authors reviewed the history of input methods used for video games, in particular previous attempts at introducing alternative input methods and how successful they have been. In [80] the author presented a set of recommendations for the more efficient use of haptic technology in computer interaction techniques for visually impaired people and those with physical disabilities. In [58] the authors proposed a situated communication environment designed to foster an immersive experience for the visually and hearing impaired. More precisely they utilized an input and output modality combination, using spoken keywords output, nonspeech sound, sign language synthesis output, haptic 3D force-feedback output, haptic 3D navigation, and sign language analysis input.

Wearable Games

Wearable games are games that employ specialized devices incorporating computer and advanced electronic technologies. During the last decade, due to the rapid progress of technology wearable devices have greatly attracted the interest of game researchers/developers. More precisely, one of the first approaches reported was in [11] in which the authors explored how computer games can be designed to maintain some of the social aspects of traditional game play, by moving computational game elements into the physical world. They constructed a mobile multiplayer game, Pirates, to illustrate how wireless and proximity-sensing technology can be integrated in the design of new game experiences. Pirates was implemented on

handheld computers connected in a wireless local area network (WLAN), allowing the players to roam a physical environment, the game arena. In [84] the authors presented an outdoor/indoor augmented reality first person application, namely the ARQuake, which was of the desktop game Quake in an attempt to investigate how to convert a desktop first person application into an outdoor/indoor mobile augmented reality application. A preliminary version of this work can be found in [21]. The player wore the wearable computer on his/her back, placed the Head Mounted Display (HMD) on his/her head, and held a simple two-button input device, a haptic gun. In [2] the authors presented an example of how the gap between virtual and physical games can be bridged using sensing technology from a wearable computer. To this end they proposed Unmasking Mister X, a game which incorporates sensor data from all the players. Each player was equipped with a sensing device and a personal digital assistant (PDA) (palmtop computer) or a head-mounted display. The game was played by walking around and approaching people to find out who is Mister X. In [35] the authors conducted an initial experiment with inexpensive body-worn gyroscopes and acceleration sensors for the Chum Kiu motion sequence in Wing Chun (a popular form of Kung Fu). In [20] the authors described the efforts on designing games for wearable computing technology taken by the 23 students of the project PEnG—Physical Environment Games. In [9] the authors used a Global Positioning System (GPS) device to extract the coordinates of the players position and create a game that explored the ability of one player competing with the others. The developed game was based on Dune 2 [83] in which the players fight for the dominance of a resource rich desert planet Dune.

Affective Evaluation of Players

Another important issue in affective gaming is the affective evaluation of players, closely related to the so-called player experience. The term player experience is quite ambiguously defined as the interaction with a game design in the performance of cognitive tasks, with a variety of emotions arising from or associated with different elements of motivation, task performance and completion, or as the structures of player interaction with the game system and with other players in the game. Its goal is to provide a motivating and fun experience for the player. Gameplay experience consists of three factors: the game quality (game system experience), the quality of player interaction (individual player experience), and the quality of this interaction in a given social, temporal, spatial or other context. Game system experience is controlled by game developers through software and game testing. Individual game experience can be assessed through psychophysiological player testing [36], eye tracking [4, 75], persona modeling, game metrics behaviour assessment [5], player modelling [86], qualitative interviews and questionnaires and Rapid Iterative Testing and Evaluation. Player context experience is assessed with ethnography, cultural debugging, playability heuristics, qualitative interviews, questionnaires and multiplayer game metrics [25]. Martinez and Yannakakis in

[55] argue that, when trying to assess which of the game levels that the player went through was more fun, interesting or less frustrating, questionnaires should *compare* game levels, instead of directly *rating* them. This approach has been shown to eliminate subjectivity across ratings from a particular player and from across different players who fill in the questionnaires and was put to use while recording the Platformer Experience Database (PED) [46], one of the few freely available datasets which combines visual, affective and game behaviour data.[1] In [25] the authors proposed an approach that formalizes the creation of evaluating methods as well as a roadmap for applying them in the context of serious games. They focused on physiological and technical metrics for game evaluation in order to design and evaluate gameplay experience. In [59] the authors extracted psychophysiological recordings of electrodermal activity (EDA) and facial muscle activity (EMG) and combined them with a Game Experience Questionnaire (GEQ) in order to measure reliably affective user experience (UX). They also introduced sound and music control in order to measure its influence on immersion, tension, competence, flow, negative affect, positive affect, and challenge. More recently, in [10] the author tried to understand engagement on the basis of the body movements of the player and connect it with the player's engagement level and affective experience.

Affective Interaction in Games

As mentioned in the previous Section, players interaction constitutes an important factor in measuring gameplay experience. To this end, various approaches have been proposed to measure the affective interaction in games. More precisely, in [54] the authors presented a method of modeling user emotional state, based on a users physiology, for users interacting with play technologies. Their modelled emotions captured usability and playbility, and exhibited the same trends as reported emotions for fun, boredom, and excitement. In [53] the authors extended their previous work in order to model emotion using physiological data. They proposed a fuzzy logic model that transformed four physiological signals into arousal and valence and a second fuzzy logic model that transformed arousal and valence into five emotional states relevant to computer game play: boredom, challenge, excitement, frustration, and fun, proposing in that way a method for quantifying emotional states continuously during a play experience. In [82] authors introduced the Koko architecture, which improved developer productivity by creating a reusable and extensible environment, yielded an enhanced user experience by enabling independently developed applications to collaborate and provided a more coherent user experience than currently possible and enabled affective communication in multiplayer and social games. The Siren game [92] utilised affective information in two ways: directly, via questionnaires filled-in by players during gameplay and

---

[1]Database available at http://institutedigitalgames.com/PED/

after each game turn, and via web cameras which estimated facial expressions. In the first case, players self-reported the perceived level of conflict, when trading virtual resources with other players; an objective of the game was to maintain perceived conflict levels between pre-set minimum and maximum values, so as to engage and not frustrate players, so this information was used to procedurally generate game quests predicted to fulfil that requirement. Similarly, facial expressions and cues (e.g. visual attention [4]) were used to estimate player involvement in the game, when associated with player behaviour (progress in the game and completing the game quests).

Existing Commercial Games

In this Section we will discuss existing commercial games that use the physiological and neurological cues measured using the sensors presented above. The commercial affective games that have been developed include the following:

- *Bionic Breakthrough* (Atari 1983), a bounce the ball into a brick wall game. The player wears a headband on his head whose sensors are supposed to pick up any facial movements or muscle twitches, in order to control the movements of the paddle and use is as input instead of an ordinary joystick.
- *Missile Command* (Atari 1980), in which the player has to destroy moving targets. The heart beat rate of a player is measured and used to change the nature of the challenge the game presents. The aim is to keep engagement within an optimum range.
- *Oshiete Your Heart* (Konami 1997), a Japanese dating game. The heart beat rate and sweat level of a player is measured. The goal is to use the measurements in order to influence the outcome of a date.
- *Zen Warriors*, a fighting game where players have to calm and concentrate in order to perform their finishing move.
- *Left 4 Dead 2* (Valve 2008) a first person shooter video game, where the player's stress level, measured as the electric response of the player's skin, determines the pace of the game. The goal is to make the game easier if the player is too stressed.
- *Nevermind* (Flying Mollusk 2015) is a horror game that lets you use biofeedback to affect gameplay and make the game scarier.
- *Journey to Wild Divine* (Wild Divine 2005) a biofeedback video game system promoting stress management and overall wellness through the use of breathing, meditation and relaxation exercises.
- *Throw Trucks With Your Mind* (Lat Ware 2013) a first-person puzzler in which players must use a combination of concentration and mental relaxation to pick up objects, and throw them at enemies, obstacles and other players.

Affective Gaming Scenarios and Challenges

In this Section we will first examine new gaming scenarios and present some of their applications. We will also elaborate on the challenges that such scenarios raise.

Affective Gaming Scenarios

The future in affective gaming lies in sensorless systems that will enable the system to extract the players' behavioural clues without employing specialized equipment. In that way the issues raised regarding the realistic immersion of the players in the gaming scenario are resolved, as the players are now free to act as they wish, not being constrained by sensors that limit their actions. Below we present three such scenarios in detail. More precisely, we begin with the special case of these scenarios that involves a group of people residing in the same space and proceed with a scenario that allows the use of such games from players with special needs. The most general case involves any kind of players that do not reside in the same space. In more detail:

- The first scenario is that of a game played among a number of players residing in the same space. The aim is to enable the players interaction and recognize human behaviour in an individual and social group level. The players under examination play with a games machine (for example Microsoft Kinect [57]). For such a scenario, a number of low cost cameras (RGB and depth) is used to create a multicamera scenario in which the entire scene (360°) is recorded. In that way, several issues raised by occlusions (either due to space limitations or caused by the co-existence of many people in the same space) and clutter are tackled as the information from different viewpoints is provided. Computer vision techniques are applied to extract and efficiently recognise each player's facial expressions, body pose and actions, in order to properly fuse them and acquire each player's emotional state. The extracted emotional state, recognized within an action context, is then combined with the behavioural patterns exhibited by the rest of the group to define the possible interactions of the groups members, as well as the relationships formed among them, exploiting simultaneously the time dynamics. The individual actions and emotional state, as well as the group interactions and relationships are used to predict future individual actions and emotional states, but also subsequent groups behaviour. Those behavioural cues are then used to alter the gameplot and offer a more realistic and satisfactory gameplay.
- The second scenario is derived from the first one, if we consider now that the players may be constricted by physical limitations, as in the case of players with special needs (for example a person being in a wheelchair). In such a case the players are not free to act as they wish, but are constrained by the use of

specialized equipment that will allow them to freely navigate in space. In such cases, wearable devices have to be included in the scenario. For example, the use of HMDs and PDAs will enable us to extract the player's position in the game, in order to model more effectively the interaction among players. Moreover, possible occlusions and clutter have to be modeled in a different way, so as to take under consideration the existence of a wearable device. Furthermore, such a device may obscure part of the face/body from the cameras. Therefore, the behavioural cues extracted from the player's visible body parts, as well as his/her actions should be emphasized to compensate for the lack of other missing input feature sources. Due to physical limitations being imposed, the actions may be restricted, thus making affect recognition especially important. Moreover, the techniques that will be used to fuse the available sources of information should be able to weight the available sources taking under consideration the constraints having been imposed by the players needs. For example, in a games scenario in which the players are in a wheelchair, thus having their actions restricted, the effect emotions have in the gameplot, gameplay and game outcome should be emphasized.

- The third scenario is the general case of the two previous scenarios, lifting all possible space limitations appearing in a game. The players in such a scenario can be of any kind (with or without special needs), while most importantly, may or may not reside in the same space. It will be possible, for example, for players to be in their own living rooms while playing. Therefore the game should be now able to construct a virtual environment in which all players will be effectively immersed and in which they will be able to freely interact with each other. 'Virtual' space limitations in terms not only of occlusions and clutter, but also of space limitations due to the use of wearable devices have to be imposed by the system. The game should be therefore able to not only recognize each players emotional state, but also combine their physical presence in a virtual world, thus reconstructing an environment in which each player's actions will affect the entire group not only emotionally, but also in a physical way. The players, although being in different for example rooms should experience the feeling of being a part of the same virtual environment, in order to ensure maximum immersion. Summarizing, the role of affect recognition in such a scenario will be of greatest importance, as the possible actions/interactions/relationships observed will be controlled by the game through the recognized emotional states.

A schematic representation of the proposed scenarios is depicted in Fig. 4.2. Summarizing, we can see that the need of incorporating affective gaming in modern control of games is crucial, not only for the simple cases (first scenario) but for more elaborated ones (second and third scenarios) in which the physical presence of a player is limited or not even required. However, typical computer vision techniques do not suffice for such applications, as several issues are raised that remain to be solved. In the next Section we elaborate on the issues raised in such scenarios in detail.
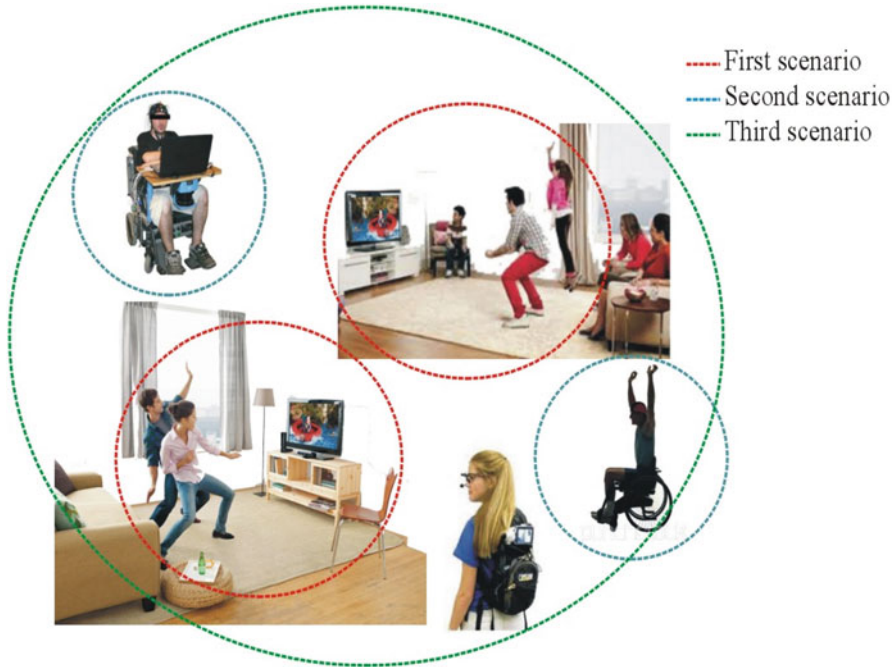
Fig. 4.2 A diagram of the three scenarios

Affective Gaming Challenges

Several issues have to be tackled in affective gaming scenarios in order to achieve a realistic interaction of the player with the game. Regarding the first and simplest scenario, in an individual level, the first step involves the real-time detection of a player and of his/her body parts. Additionally, the problem of recognizing his/her actions and emotions has been widely studied in the past, but involving a set of predefined classes under examination. Therefore, the introduction of spontaneity, as the player may express himself/herself in any way that he/she wishes, constitutes an extra challenge. Moreover, the proposed scenario employs many cameras so as to extract information from different viewpoints. Although the information from different viewpoints aids in correctly detecting the player and recognizing his body pose and emotions, finding efficient methods to fuse those pieces of information remains an open problem. In a group level, the goal is to extract the social group behaviour. Several extra challenges exist, for example several occlusions due to space limitations but also to the presence of many people in the same space. The free interaction among the players and the way that affects their subsequent emotions and states is also a novel field of research. After having efficiently recognized the emotional state of each individual player within an action context as performed in the single-player game scenario, the next step is to study the players as members of

a social group, that is to identify the way each player interacts with each other and also to identify relationships built within this interaction framework. For example, do a player's actions reveal a friendly/aggressive mood towards the other players? Can we use the player's actions to predict subsequent actions? When it comes to the whole group, do players choose a leader, even subconsciously, whom they tend to mimic? Are cohesion (i.e. tendency of people to form groups) relationships formed? How are all of the aforementioned interactions and relationships developed in time? Can we exploit the information their dynamics have to offer?

Regarding the second scenario, several issues are raised by the use of wearable devices. First, the wearable device can obscure a significant part of the face/body. Moreover, its presence may lead the player to exhibit unusual behavioural parts due to the, even subconscious, limitations imposed by the use of the device. This also affects the interaction among players and of course their subsequent actions. Therefore novel methods to model and predict those behavioural patterns have to be proposed.

The problem becomes more complicated for the third scenario, in which space limitations are eliminated. The players may not now be limited by the presence of other players, since they may not reside in the same space. The game however should be able to 'combine' the presence of multiple players in one 'virtual' environment, in which their interaction will be possible. And of course, this has to be performed in a realistic way, so as to ensure maximum immersion to the game, offering at the same time a better gameplay experience.

Summarizing, as we can see, such scenarios involve many interdisciplinary fields. Besides the obvious computer vision techniques that have to be employed in order to extract behavioural cues, input from psychologists has to be provided. More precisely, input from psychologists is required in order to properly define the scenario under examination in terms of the emotions and actions that are more likely to be observed during the game (emotions and personality traits). Which is the role that the expressed emotions play in the overall gameplay experience? How realistic should the expressed emotions be in order to maintain player engagement? The pool of possible interactions among the players as well as the relationships that they are most likely to form while in the game should also be defined. Which modalities (speech, gestures, facial expressions) should be used and to which should the game emphasize? How should the gameplay be adopted to the players affective states? Input from psychologists is also required to extract the ground truth concerning the emotional state of each player and to explain the way it affects his/her actions as well his/her interactions with other players and the relationships built among them as member of a social group. The challenging nature of the proposed scenarios regarding behaviour understanding, in combination with the scarcity of available datasets, constitutes the proposed research a novel field, even for psychologists.

Applications of Affective Games

Creating games that will be able to understand the players' emotional states and actions will enable the creation of games in which the progression of the gameplot will correspond to the players needs, thus ensuring more realistic and satisfactory gameplay experience. More versatile ways of human computer interaction will be developed and more player-friendly applications will be addressed. Indeed, such gaming scenarios can be widely applied to many fields leading to applications that greatly enhance the interaction of the players with the games. More specifically some possible applications include:

- *Serious games*, that is games with an initial purpose other than pure entertainment, usually used to teach something to the player. They are especially useful in helping younger players develop collaborative skills, while engaging their focus. They are used to teach math, language, science, social skills [92], etc. Moreover, they can be massively used in "virtual" universities, providing education through electronic media (typically the Internet). They offer flexibility to students that cannot attend physical courses due to distance or require flexible time schedules. This type of games also includes training games. A typical example is the earthquake simulation games, in which the player learns how to safely evacuate a building in the case of an earthquake [87].
- *Multimedia annotation* via games, enabling players to implicit tag multimedia content, aiming at providing fast and accurate data retrieval. The annotation of multimedia data can be performed using either the typical way (text tags) of by using the players nonverbal reactions (e.g., facial expressions like smiles or head gestures like shakes, laughter when seeing a funny video etc.) [64].
- *Entertainment*, that is for the production of interactive applications or virtual worlds, in which realistic avatars exist. These can be used not only in games, but also for movies production. The entertainment industry is greatly affected by games and vice versa.

Conclusions

The existing games scenarios seem to have undergone a major transformation through the past 5 years, due to the recent technological advances that allow for robust, sensorless and real-time interaction of the players with the game. Indeed old-fashioned games required from the player to use a specialized input device in order to interact with the game. The player had a non-existent, in reality, feeling of controlling the game, even though the game plot and game responses to him/her were predefined. In order to create more realistic games in which the player's emotional state/actions/needs would be used to progress the gameplot and alter the gameplay experience accordingly, several more elaborated methods were proposed. The interest of the scientific community has been shifted towards affective gaming

during the last few years, as the incorporation of affect recognition in games scenarios allowed for a more realistic gameplay experience for the players. In this paper we elaborated on the existing approaches regarding affective computing and discussed the recent technological advances that progressed the field. We reviewed the different sources of acquiring affect information and investigated issues that arise in affective gaming scenarios, such as the affective evaluation of players and the affective interaction in games. We presented the e existing commercial affective gaming applications and introduced new gaming scenarios. Last, we discussed about the challenges that affective gaming scenarios have to tackle in order to achieve a more realistic gameplay experience.

References

1. Adolphs R (2002) Recognizing emotion from facial expressions: psychological and neurological mechanisms. Behav Cogn Neurosci Rev 1:21–62
2. Antifakos S, Schiele B (2002) Bridging the gap between virtual and physical games using wearable sensors. In: Sixth international symposium on wearable computers, Seattle
3. Argyle M (1988) Bodily communication
4. Asteriadis S, Karpouzis K, Kollias S (2014) Visual focus of attention in non-calibrated environments using gaze estimation. Int J Comput Vis 107(3):293–316
5. Asteriadis S, Karpouzis K, Shaker N, Yannakakis GN (2012) Towards detecting clusters of players using visual and gameplay behavioral cues. Proc Comput Sci 15:140–147
6. Asteriadis S, Shaker N, Karpouzis K, Yannakakis GN (2012) Towards player's affective and behavioral visual cues as drives to game adaptation. In: LREC workshop on multimodal corpora for machine learning, Istanbul
7. Bateman C, Nacke L (2010) The neurobiology of play. In: Proceedings of future play 2010, Vancouver, pp 1–8
8. Bernhardt D, Robinson P (2007) Detecting affect from non-stylised body motions. In: International conference on affective computing and intelligent interaction, Lisbon
9. Bertelsmeyer C, Koch E, Schirm AH (2006) A new approach on wearable game design and its evaluation. In: Proceedings of 5th ACM SIGCOMM workshop on network and system support for games, Singapore
10. Bianchi-Berthouze N (2010) Does body movement affect the player engagement experience? In: International conference on Kansei engineering and emotion research, Paris
11. Björk S, Falk J, Hansson R, Ljungstrand P (2001) Pirates: using the physical world as a game board. In: Proceedings of interact, Tokyo
12. Blair RJR, Morris JS, Frith CD, Perrett DI, Dolan RJ (1999) Dissociable neural responses to facial expressions of sadness and anger. Brain 122:883–893

13. Borod JC, Obler LK, Erhan HM, Grunwald IS, Cicero BA, Welkowitz J, Santschi C, Agosti RM, Whalen JR (1998) Right hemisphere emotional perception: evidence across multiple channels. Neuropsychology 12:446–458
14. Bruce V, Young AW (1986) Understanding face recognition. Br J Psychol 77:305–327
15. Calder AJ, Burton AM, Miller P, Young AW, Akamatsu S (2001) A principal component analysis of facial expressions. Vis Res 41:1179–1208
16. Camurri A, Mazzarino B, Ricchetti M, Timmers R, Timmers G (2004) Multimodal analysis of expressive gesture in music and dance performances. In: Gesture-based communication on human-computer interaction. Springer, Berlin/New York
17. Caporusso N, Mkrtchyan L, Badia L (2010) A multimodal interface device for online board games designed for sight-impaired people. IEEE Trans Inf Technol Biomed 14:248–254
18. Caridakis G, Karpouzis K, Wallace M, Kessous L, Amir N (2010) Multimodal user's affective state analysis in naturalistic interaction. J Multimodal User Interfaces 3(1):49–66
19. Caridakis G, Wagner J, Raouzaiou A, Curto Z, Andre E, Kostas K (2010) A multimodal corpus for gesture expressivity analysis. In: Multimodal corpora: advances in capturing, coding and analyzing multimodality, Valletta
20. Cinaz B, Dselder E, Iben H, Koch E, Kenn H (2006) Wearable games—an approach for defining design principles. In: Student colloquium at the international symposium on wearable computers ISWC06, At Montreux
21. Close B, Donoghue J, Squires J, Bondi PD, Morris M, Piekarski W (2000) Arquake: an outdoor/indoor augmented reality first person application. In: 4th international symposium on wearable computers, Atlanta
22. Cohn J, Schmidt K (2004) The timing of facial motion in posed and spontaneous smiles. Int J Wavelets Multiresolut Inf Process 2:121–132
23. Cowie R, Douglas-Cowie E, Karpouzis K, Caridakis G, Wallace M, Kollias S (2008) Recognition of emotional states in natural human-computer interaction. In: Tzovaras D (ed) Multimodal user interfaces. Springer, Berlin/Heidelberg, pp 119–153
24. DeGroot D, Broekens J (2003) Using negative emotions to impair game play. In: 15th Belgian-Dutch conference on artificial intelligence, Nijmegen
25. Drachen A, Gbel S (2010) Methods for evaluating gameplay experience in a serious gaming context. Int J Comput Sci Sport 9:40–51
26. Ekman P, Davidson RJ (1994) The nature of emotion: fundamental questions. Oxford University Press, New York
27. Ekman P, Friesen W (1969) Nonverbal leakage and clues to deception. Psychiatry 32(1):88–105
28. Ekman P, Friesen W (1974) Detecting deception from the body or face. Personal Soc Psychol 29(3):288–298
29. Faust M, Yoo Y (2006) Haptic feedback in pervasive games. In: Third international workshop on pervasive gaming applications, PerGames, Dublin
30. Gauthier I, Tarr MJ, Aanderson A, Skudlarski P, Gore JC (1999) Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat Neurosci 2:568–573
31. Gilleade K, Dix A, Allanson J (2005) Affective videogames and modes of affective gaming: assist me, challenge me, emote me. In: DiGRA, Vancouver
32. Gunes H, Piccardi M (2009) Automatic temporal segment detection and affect recognition from face and body display. IEEE Trans Syst Man Cybern B Spec Issue Hum Comput 39(1):64–84
33. Hasselmo ME, Rolls ET, Baylis GC (1989) The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. Behav Brain Res 32:203–218
34. van Heijnsbergen CCRJ, Meeren HKM, Grezes J, de Gelder B (2007) Rapid detection of fear in body expressions, an ERP study. Brain Res 1186:233–241
35. Heinz EA, Kunze KS, Gruber M, Bannach D, Lukowicz P (2006) Using wearable sensors for real-time recognition tasks in games of martial arts—an initial experiment. In: IEEE symposium on computational intelligence and games, Reno/Lake Tahoe

36. Holmgård C, Yannakakis GN, Martínez HP, Karstoft KI, Andersen HS (2015) Multimodal ptsd characterization via the startlemart game. J Multimodal User Interfaces 9(1):3–15
37. Hossain SKA, Rahman ASMM, Saddik AE (2010) Haptic based emotional communication system in second life. In: IEEE international symposium on haptic audio-visual environments and games (HAVE), Phoenix
38. Hudlicka E (2003) To feel or not to feel: the role of affect in humancomputer interaction. Int J Hum-Comput Stud 59:1–32
39. Hudlicka E (2008) Affective computing for game design. In: Proceedings of the 4th international North American conference on intelligent games and simulation (GAMEON-NA), Montreal
40. Hudlicka E (2009) Affective game engines: motivation and requirements. In: Proceedings of the 4th international conference on foundations of digital games, Orlando
41. Hudlicka E, Broekens J (2009) Foundations for modelling emotions in game characters: modelling emotion effects on cognition. In: Affective Computing and Intelligent Interaction and Workshops (ACII), 3rd international conference on, pp 1–6, IEEE
42. Hudlicka E, McNeese MD (2002) Assessment of user affective and belief states for interface adaptation: application to an air force pilot task. User Model User-Adapt Interact 12(1):1–47
43. Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17(11):4302–4311
44. Kapoor A, Burleson W, Picard RW (2007) Automatic prediction of frustration. Int J Hum-Comput Stud 65:724–736
45. Kapoor A, Picard RW, Ivanov Y (2004) Probabilistic combination of multiple modalities to detect interest. In: International conference on pattern recognition, Cambridge
46. Karpouzis K, Shaker N, Yannakakis G, Asteriadis S (2015) The platformer experience dataset. In: 6th international conference on affective computing and intelligent interaction (ACII 2015) conference, Xi'an, 21–24 Sept 2015
47. Kleinsmith A, Bianchi-Berthouze N, Steed A: Automatic recognition of non-acted affective postures. IEEE Trans Syst Man Cybern B: Cybern 41(4):1027–1038 (2011)
48. Kleinsmith A, Silva RD, Bianchi-Berthouze N (2006) Cross-cultural differences in recognizing affect from body posture. Interact Comput 18:1371–1389
49. Kwon YC, Lee WH (2010) A study on multi-touch interface for game. In: 3rd international conference on human-centric computing (HumanCom), Cebu
50. Leite I, Martinho C, Pereira A, Paiva A (2008) iCat: an affective game buddy based on anticipatory mechanisms. In: Proceedings of the 7th international joint conference on autonomous agents and multiagent systems, Estoril
51. Lyons MJ, Budynek J, Akamatsu S (1999) Automatic classification of single facial images. IEEE Trans Pattern Anal Mach Intell 21(12):1357–1362
52. Ma Y, Paterson H, Pollick FE (2006) A motion capture library for the study of identity, gender, and emotion perception from biological motion. Behav Res Methods 38:134–141
53. Mandryk R, Atkins MS (2007) A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. Int J Hum-Comput Stud 65(4):329–347
54. Mandryk RL, Atkins MS, Inkpen KM (2006) A continuous and objective evaluation of emotional experience with interactive play environments. In: Proceedings of the SIGCHI conference on human factors in computing systems, Montréal
55. Martinez HP, Yannakakis GN, Hallam J (2014) Don't classify ratings of affect; rank them! IEEE Trans Affect Comput 5(3):314–326
56. Mehrabian A, Friar J (1969) Encoding of attitude by a seated communicator via posture and position cues. Consult Clin Psychol 33:330–336
57. Microsoft (2010) Xbox kinect. http://www.xbox.com/en-GB/kinect
58. Moustakas K, Tzovaras D, Dybkjaer L, Bernsen N, Aran O (2011) Using modality replacement to facilitate communication between visually and hearing-impaired people. IEEE Trans Multimed 18(2):26–37

59. Nacke LE, Grimshaw NM, Lindley ACA (2010) More than a feeling: measurement of sonic user experience and psychophysiology in a first-person shooter game. Interact Comput 22:336–343
60. Nacke LE, Mandryk RL (2010) Designing affective games with physiological input. In: Workshop on multiuser and social biosignal adaptive games and playful applications in fun and games conference (BioS-Play), Leuven
61. Nusseck M, Cunningham DW, Wallraven C, Bülthoff HH (2008) The contribution of different facial regions to the recognition of conversational expressions. J Vis 8:1–23
62. Ouhyoung M, Tsai WN, Tsai MC, Wu JR, Huang CH, Yang TJ (1995) A low-cost force feedback joystick and its use in PC video games. IEEE Trans Consum Electron 41:787
63. Pantic M, Rothkrantz LJM (2003) Toward an affect-sensitive multimodal human-computer interaction. Proc IEEE 91:1370–1390
64. Pantic M, Vinciarelli A (2009) Implicit human centered tagging. IEEE Signal Process Mag 26:173–180
65. Paolis LD, Pulimeno M, Aloisio G (2007) The simulation of a billiard game using a haptic interface. In: IEEE international symposium on distributed simulation and real-time applications, Chania
66. Park W, Kim L, Cho H, Park S (2009) Design of haptic interface for brickout game. In: IEEE international workshop on haptic audio visual environments and games, Lecco
67. Picard R (2000) Towards computers that recognize and respond to user emotion. IBM Syst J 39:705
68. Pollak SD, Messner M, Kistler DJ, Cohn JF (2009) Development of perceptual expertise in emotion recognition. Cognition 110:242–247
69. Russell JA (1980) A circumplex model of affect. J Personal Soc Psychol 39:1161–1178
70. Saari T, Ravaja N, Laarni J, Kallinen K, Turpeinen M (2004) Towards emotionally adapted games. In: Proceedings of presence, Valencia
71. Salah AA, Sebe N, Gevers T (2010) Communication and automatic interpretation of affect from facial expressions. In: Gokcay D, Yildirim G (eds) Affective computing and interaction: psychological, cognitive, and neuroscientific perspectives, IGI Global, pp 157–183
72. Sanghvi J, Castellano G, Leite I, Pereira A, McOwan PW, Paiva A (2011) Automatic analysis of affective postures and body motion to detect engagement with a game companion. In: HRI'11, Lausanne, pp 305–312
73. Savva N, Bianchi-Berthouze N (2012) Automatic recognition of affective body movement in a video game scenario. In: International conference on intelligent technologies for interactive entertainment, Genova
74. Savva N, Scarinzi A, Bianchi-Berthouze N (2012) Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience. IEEE Trans Comput Intell AI Games 4(3):199–212
75. Shaker N, Asteriadis S, Yannakakis GN, Karpouzis K (2013) Fusing visual and behavioral cues for modeling user experience in games. IEEE Trans Cybern 43(6):1519–1531
76. Silva RD, Bianchi-Berthouze N (2004) Modeling human affective postures: an information theoretic characterization of posture features. Comput Animat Virtual Worlds 15:269–276
77. Silva PRD, Madurapperuma AP, Marasinghe A, Osano M (2006) A multi-agent based interactive system towards child's emotion performances quantified through affective body gestures. In: ICPR (1)'06, Hong Kong, pp 1236–1239
78. Silva PRD, Osano M, Marasinghe A, Madurapperuma AP (2006) Towards recognizing emotion with affective dimensions through body gestures. In: International conference on automatic face and gesture recognition, Southampton
79. Sivak M, Unluhisarcikli O, Weinberg B, Melman-Harari A, Bonate P, Mavroidis C (2010) Haptic system for hand rehabilitation integrating an interactive game with an advanced robotic device. In: IEEE haptics symposium, Waltham
80. Sjöström C (2001) Using haptics in computer interfaces for blind people. In: CHI '01 extended abstracts on human factors in computing systems. ACM, New York

81. Sjöström C, Rassmus-Gröhn K (1999) The sense of touch provides new interaction techniques for disabled people. Technol Disabil 10:45–52
82. Sollenberger DJ, Singh MP (2009) Architecture for affective social games. In: Dignum F, Silverman B, Bradshaw J, van Doesburg W (eds) Proceedings of the first international workshop on agents for games and simulations, vol 5920 of LNAI. Springer, Berlin, pp 135–154
83. Studios W: Dune 2—battle for arrakis (1993). http://www.ea.com/official/cc/firstdecade/us
84. Thomas B, Close B, Donoghue J, Squires J, Bondi PD, Morris M, Piekarski W (2002) First person indoor/outdoor augmented reality application: ARQuake. Pers Ubiquitous Comput Arch 6:75–86
85. Thorpe A, Ma M, Oikonomou A (2011) History and alternative game input methods. In: International conference on computer games, Louisville
86. Togelius J, Shaker N, Yannakakis GN (2013) Active player modelling. arXiv preprint arXiv:1312.2936
87. Ververidis D, Kotsia I, Kotropoulos C, Pitas I (2008) Multi-modal emotion-related data collection within a virtual earthquake emulator. In: 6th language resources and evaluation conference (LREC), Marrakech
88. Wallbott HG, Scherer KR (1986) Cues and channels in emotion recognition. Personal Soc Psychol 51(4):660–699
89. Wong CY, Chu K, Khong CW, Lim TY (2010) Evaluating playability on haptic user interface for mobile gaming. In: International symposium in information technology (ITSim), Kuala Lumpur
90. Yang CY, Lo YS, Liu CT (2010) Developing an interactive dental casting educational game. In: 3rd IEEE international conference on computer science and information technology (ICCSIT), Chengdu
91. Yannakakis GN, Togelius J (2011) Experience-driven procedural content generation. IEEE Trans Affect Comput 2(3):147–161
92. Yannakakis GN, Togelius J, Khaled R, Jhala A, Karpouzis K, Paiva, A, Vasalou A (2010) Siren: towards adaptive serious games for teaching conflict resolution. In: 4th European conference on games based learning (ECGBL10), Copenhagen
93. Zeng Z, Pantic M, Roisman G, Huang T (2009) A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Trans Pattern Anal Mach Intell 31(1):39–58

Chapter 5
Emotion Modelling via Speech Content and Prosody: In Computer Games and Elsewhere

Björn Schuller

Abstract The chapter describes a typical modern speech emotion recognition engine as can be used to enhance computer games' or other technical systems' emotional intelligence. Acquisition of human affect via the spoken content and its prosody and further acoustic features is highlighted. Features for both of these information streams are shortly discussed along chunking of the stream. Decision making with and without training data is presented, each. A particular focus is then laid on autonomous learning and adaptation methods as well as the required calculation of confidence measures. Practical aspects include the encoding of the information, distribution of the processing, and available toolkits. Benchmark performances are given by typical competitive challenges in the field.

Introduction

The automatic recognition of emotion in speech dates back some twenty years by now looking back at the very first attempts, cf. e.g., [9]. It is the aim of this chapter to give a general glance 'under the hud' how today's engines work. First, a very brief overview on modelling of emotion is given. A special focus is then laid on speech emotion recognition in computer games owing to the context of this book. Finally, the structure of the remaining chapter is provided aiming at familiarising the reader with the general principles of current engines and their abilities, principles, and necessities.

Emotion Modelling

A number of different representation forms have been evaluated, with the most popular ones being discrete *emotion classes* such as 'anger', 'joy', or 'neutral' – usually reaching from two to roughly a dozen [51] depending on the

B. Schuller (✉)
Imperial College London, 180 Queen's Gate, SW7 2AZ London, UK
e-mail: bjoern.schuller@imperial.ac.uk

application of interest –, and a representation by continuous emotion 'primitives' in the sense of a number of (quasi-)value-continuous *dimensions* such as arousal/activation, valence/positivity/sentiment, dominance/power/potency, expectation/surprise/novelty, or intensity [43]. In a space spanned by these axes, the classes can be assigned as points or regions, thus allowing for a 'translation' between these two representation forms. Other popular approaches include *tagging* by allowing several class labels per instance of analysis (in case of two, the name *complex emotions* has been used), and calculating scores per each emotion class leading to '*soft emotion profiles*' [32] – potentially with a minimum threshold to be exceeded.

Besides choosing such a representation of emotion, one has to choose a temporal segmentation from, as the speech needs to be segmented into units of analysis. This analysis itself can be based on the spoken content or the 'way of speaking' it in the sense of prosody, articulation, and voice quality as the acoustic properties. As these two information streams tend to benefit from different temporal levels, one can choose different lengths of these units, accordingly.

Emotion and Games

It is generally believed that taking players' emotion in computer games into account possesses great potential for improved game-play and human behaviour feedback, in particular in 'serious games' [20, 46]. However, few games up to this point make use of the ability to recognise emotion from speech. This is likely owing to the fact that (1) games are not often controlled by speech, yet – however, a number of games already use this option and notably the Microsoft Xbox One console is offering speech control in many games via its Kinect interface; and (2) if no headset is used, games' and environmental audio can pose additional challenges to the use of this modality as compared, e.g., the use of bio-sensors.

Besides the commercial game "Truth or Lies", some examples of games that consider affect and emotion recognition in speech in research prototypes include [24, 29], and in particular several games for children [62, 63], especially with autism spectrum condition [35, 48] – often centred around emotion as a concept itself. Related to games, affect and emotion recognition has also been considered in educational software [26, 27]. Interestingly, also the emotion of the voice of the virtual characters of a game has been recognised automatically [38] showing that this is feasible as well. Games have, however, also been used to elicit emotion when collecting speech data 'coloured' by affect for training and evaluation of models [22, 23, 41, 53]. This makes it evident that, there is indeed 'a lot of emotion' in computer gaming that should best be captured by the game engine.

Further, a number of games exploit other modalities such as physiological measurement for emotion recognition, e.g., the "Journey to the Wild Divine" as a commercial example; other examples include mostly research prototypes [5, 19, 24, 34]. Further modalities used for emotion recognition in games include touch interaction [17], and facial expression and body gestures [48].

The remainder of this chapter is structured as follows: In section "Speech Content", we will be dealing with the recognition of emotion from the speech content, i.e., from the spoken words; then, in section "Prosodic and Acoustic Modelling" we will be considering the prosody and acoustics of the spoken words to the same end. In both of these sections, we will follow the typical chain of processing in pattern recognition systems by going from the denoising preprocessing to the feature extraction – both aiming at reaching a canonical representation of the problem of interest, usually coming at a reduction of information. Then, based on the reached representation, a decision can be made either without 'training' of a model from data, e.g., by rules such as "IF the words 'fun', 'sun', (...) are used" or "IF pitch increases and speed increases, THEN assume joy". Alternatively, as decisions are often much more complex, machine learning algorithms such as support vector machines (SVM) or neural networks (NNs) are the typically seen solutions. In section "Integration and Embedding", we will deal with practical issues when it comes to embedding 'speech emotion recognition' (SER) in an application, software system, or more specifically computer game. This includes the fusion of textual and acoustic cues, available tools ready to be used ready to 'plug and play', benchmarks on popular databases, distributed processing, confidence measures, adaptation and self-learning, and encoding of the emotion information. Then, to conclude, a summary will be given naming some existing 'white spots' in the literature that should best be filled sooner rather than later if SER is to be used widely in computer games and other real-world products.

Speech Content

We tend to express our emotion in the words we choose. Not only is this true for the words themselves, but also for their 'part of speech' (POS) classes such as noun, verb or adjective. The latter means that depending on the emotion, different frequencies of the POS classes can be observed. In fact, also semantically higher 'tagging' of words such as 'first person pronoun' or 'about religion' (such as 'Christ', 'Jesus', 'holy', 'Lord', etc.) may reveal the affect 'of the moment'. At the same time, the level of control by an individual for regulation or masking of this type of affective information is usually high, i.e., we are trained to choose these words different from our emotion in case we want to hide our feelings. In the following, let us take a look on how to model emotion in such a way that the described speech content can be exploited.

Speech Recognition and Emotion

One bottleneck in exploiting spoken or written words for the recognition of emotion of players or in general is that these need to be recognised in the first place. Even if

they should be typed rather than spoken, such as by a keybord or a game console, some preprocessing is usually necessary to 'correct' misspelling, dialects or remove or handle special characters, while ensuring not to lose relevant information such as 'smileys' as ":)" or ":-(" and alike. This is often handled by allowing for a certain Levenshtein distance (the minimum number of operations out of insertion, deletion, and substitution to match the observed word with one known by the computer – usually calculated by dynamic programming) or look-up dictionaries for different spellings. Luckily, automatic speech recognition (ASR) has matured impressively in recent time to the degree where it is available on even very low resource devices, smart phones, game consoles such as the XBox One via its Kinect, or even modern TV sets such as the Samsung UE46F7090 and alike. These ASR engines are, however, mostly focused on the recognition of (a few) keywords at a time depending on the current context. This comes, as robustness is particularly required as when gaming at home, the sounds and music as well as non-player-character (NPC) voices interfere; when playing on mobile devices, ambient noises and ever-changing reverberation patterns have to be coped with. This allows to either use an additional continuous speech larg(er) vocabulary ASR engine alongside – even if at lower requirements in terms of robustness – or to limit the emotion analysis to a few affective keywords. Surprisingly, it seems as if even lower accuracies of the ASR engine can be sufficient to be able to capture the emotion of the words [30] as long as the affective keywords are not lost or 'changed' to another affective context.

Besides the recognition of verbal units, also non-verbal units such as laughter or sighing and hesitations bear information on the emotion of a speaker. Their recognition can be handled within the ASR process, but is often executed individually, such as in [6]. Independent of that, the information of verbal and non-verbal tokens can be handled commonly such as in "*this is really <laughs> funny!*".

Textual Features

In order to represent text in a numerical representation form, there is a need for segmentation of the word or text stream at first, potentially followed by re-tagging of the words, and an optional representation in a vector space. The last choice depends on the type of decision making.

Tokenisation and Tagging

A first step is – similar to the processing of an acoustic speech signal as described later – the *tokenisation* in the sense of a segmentation. Typical approaches to this end include sequences of *N* characters or words, the latter usually delimited by spaces or special characters when considering written text. Such sequences are known as *N*-grams, or character *N*-grams and word *N*-grams, respectively. Typical lengths of sequences are roughly three to eight characters or one to three words [47].

Either before or after this tokenisation, a (re-)*tagging* of the tokens can take place. Above, POS classes and higher semantical classes were named already. Besides, stemming is a popular way of re-tagging tokens in order to cluster these by their word stem. An example would be "stems", "stemming", or "stemmed" re-tagged as "stemX". This increases the number of observations of words of the same stem and likewise allows to train more meaningful models as long as the different morphological variants indeed do not represent different emotional 'colourings'. Differently tagged variants of the text can be combined, to combine different levels of information and highlight different aspects, such as the original 'fine grained' words (usually in the order of some thousand different entries) alongside their semantic classes (rather in the order of a few hundreds) alongside their POS classes (rather some tenths).

Vector Space Modelling

In order to reach a feature representation in terms of a vector of numerical variables that allows, e.g., for the calculation of distances between such vectors belonging to different phrases, a popular approach is to count the frequency of occurrence of each token in the text of analysis. This is known as term frequency modelling [21]. Then, a feature vector is constructed by using one such feature per entry in the 'vocabulary' of known different tokens. As most texts of analysis will consist of very few words when dealing with analysis of spoken language in search of affective cues, the feature vector will accordingly contain mostly zero values. Different variants exist to normalise such a term frequency feature vector, such as using the logarithmic term frequency, dividing by the number of tokens in the text of analysis or by the number of occurrences as seen in the training material. However, it has repeatedly been observed that, given a large database for training of a machine learning algorithm, the effect of different such 'normalisation approaches' is comparably minor [47], and even a simple boolean representation of the term frequency can be efficient and sufficient when dealing with short phrases.

Zero-Resource Modelling

Without data to train from, emotion can be recognised from text thanks to a richer number of available affective dictionaries and other related knowledge sources such as ConceptNet, General Inquirer, or WordNet [31, 47]. Usually, one 'looks up' the words in the text of analysis in the according knowledge resource and computes a score per emotion or a value for an affect dimension according to the entries in these knowledge sources. Alternatively, one computes the distance in the resource to the affective tag(s) of interest, e.g., the distance in a word relation database between "sun" and "happniess" or alike.

Learning

Alternatively, one can train standard machine learning algorithms on annotated text. Most popular solutions include SVMs due to their ability to handle large feature spaces. Depending on the feature representation of the text, these can have several thousand entries.

Prosodic and Acoustic Modelling

Let us now have a look at the speech signal by exploiting 'how' something is being said.

Speaker Separation and Denoising

Considerable progress had been made over the last years in (blind) audio source separation allowing for high(er) quality separation of a speaker even in adverse acoustic conditions or even in the presence of other overlapping speakers. Some interesting approaches include a range of variants of Non-Negative Matrix Factorisation (NMF) [59] or the usage of (recurrent) neural networks (preferably with memory [60]) to estimate clean from noisy speech or at least 'clean' features derived from noisy speech. In the luxurious case of availability of several microphones such as in a microphone array, e.g., four in the case of Microsoft's Kinect sensor, Independent Component Analysis (ICA) or derivatives can separate as many sources as microphones with a very high quality. One can also mix single- and multi-microphone approaches, e.g., by ICA-NMF hybrids. Most of these approaches target additive noise, however, also convolutional artifacts such as by reverberation due to a changing room-impulse response in a mobile setting can be handled – again, data-driven approaches have lately become very powerful in this respect [60]. Surprisingly few works have investigated the usage of speaker separation for the recognition of emotion (e.g., [59]) from speech as compared to those works considering it for ASR. Interestingly, such separation also allows for recognition of group emotion.

Prosodic and Acoustic Features

Prosody is usually summarised by three feature groups: intensity, intonation, and 'rhythm'. Whereas in emotional speech synthesis, these are mostly focussed upon, in the recognition, also further acoustic features are of crucial relevance. In terms of being most descriptive, voice quality is to mention next. This includes the harmonics to noise ratio, the jitter and shimmer – micro pertubations of pitch and energy,

and other aspects describing the quality of the voice such as breathy, whispering, shouting, etc. Finally, a rich selection of further spectral and cepstral features is usually considered, such as spectral band energies and ratios – often presented in a manner closer to human hearing sensation such as mel frequency and auditory spectrum coefficients – or formants, i.e., spectral maxima described by position, amplitude, and bandwidth – usually in the order of the first five to seven.

The above named features are usually computed per 'frame' after windowing the speech signal with a suited function such as a rectangular function for analysis in the time domain and a smooth function such as Hamming or Hanning (basically, the positive half wave of a cosine signal) or Gaussian (allowing to come closest to the Heisenberg-alike time-frequency uncertainty resolution optimum). A typical frame length then is around 20–50 ms, with a frame shift of around 10 ms resembling a feature sampling frequency of 100 frames per second (fps) at this so called 'low level' – overall one thus speaks of low-level descriptors (LLDs) at this point. To add to the feature basis, further LLDs can be derived from the original contours such as by calculating delta regression or correlation coefficients. Based on the LLDs, statistical functionals are then mostly applied to lead to a supra-segmental view – emotion is a usually better modelled at a feature-level for a window of analysis of around 1 s [45] depending on the LLD type. These functionals project a LLD onto a single value per 'macro' time frame (fixed length or semantically or syntactically meaningful such as per word or phrase). Typical functionals include the lower order moments such as mean, standard deviation, kurtosis, skewness, or extrema, such as minimum, maximum, range, percentiles, segments, or even spectral functionals. This process can be carried out systematically leading to a brute-force feature extraction: First, LLDs are extracted – then, per LLD derivations are produced. Next, for each of the overall LLDs – be it original or derived – the same functionals are computed – potentially even hierarchically such as the mean of maxima or the maxima of means. This can also be considered across different window lengths or cross LLD, thus easily leading to a feature basis of several thousand features. Usually, one then carries out a data-driven feature selection to reduce the space to a few hundred or even less features. Over the years, sufficient experience was gained to formulate recommendations for such compact feature sets. However, as the emotion representation and type targeted can vary considerably, it may be worth to adapt the feature set to a specific task or domain and culture (cf. also below). Also, one can execute a controlled brute-forcing, thus not necessarily producing all possible LLD/functional/(functional) combinations (which can sometimes be meaningless as the minimum pitch value which is often simply zero), but rather searching in the space of possible combinations in a targeted way by combining expert knowledge and brute-forcing efficiently [33].

More recently, unsupervised learning of features has been considered as an alternative of expert-crafted or brute-forced descriptors. This comes often in combination with 'deep learning' approaches [25, 57] based on (sparse) auto-encoders or similar. An interesting 'conventional' method for unsupervised acoustic feature generation is the Bag-of-Audio-Words variant. It is similar to the Bag-of-Words modelling described above for linguistic content features. However, it produces

the 'audio words' in unsupervised manner by vector quantisation over some (often raw) acoustic representation such as in the spectrum, e.g., by k-means clustering or similar. Then, frequency of occurrences of these audio words are counted.

Zero-Resource Modelling

As opposed to the linguistic exploitation of the spoken content, there exists practically no acoustic SER engine that is based on (human-readable) rules that have not been learnt from data by a machine learning algorithm. This does, however, not mean that it is not possible to implement what is known from the literature on how speech changes depending on the emotion, such as in [40]. It rather means that usually, the task is of non-linear and complex nature and a few rules will not suffice. However, the advantage of not needing training data seems obvious: such a model will not be biased by the chracteristics of the speakers contained in the training data – may it be their gender, age, social class, origin, culture, or simply their co-influencing states and traits at the time of the recording of the training data. For example, in serious gaming for specific target groups such as young individuals with Autism Spectrum Condition [48], training data may be so rare and the variation of the deviation from the 'typical' so large that it can be more promising to rely on a few basic rules.

An interesting alternative version in case of limited availability of (labelled) human speech samples is the synthesis of training material [52]. Given a high quality speech synthesiser that can produce speech in different emotions, one can produce sheer infinite amounts of training material for an emotion recogniser. As such a speech synthesiser is often given in computer games, this may be a promising road in affect recognition for gaming or dialogue systems. Obviously, one can combine synthesised speech with human speech in training of classifiers often leading to an additional gain in performance.

Learning

There are at best trends visible which machine learning algorithms are preferred in the literature, as in principle, both, 'static' algorithms (with a fixed feature dimension and stream/time series length) as well as 'dynamic' algorithms (that can cope with time series of unknown length) can be used [50]. This depends on the feature representation of choice: the functional representation is usually modelled by static approaches, whereas the 'raw' LLDs are best modelled by the dynamic alternative [50]. Popular static approaches include SVMs, NNs, and decision trees, but also more simple solutions such as k-Nearest Neighbours (kNN) and related distance-based solutions, or naive Bayes and related simple statistical considerations. As for static approaches, Hidden Markov Models (HMMs) and more

general Dynamic Bayesian Networks or Graphical Models prevail over distance-based alternatives such as Dynamic Time Warping (DTW). Noteworthy, recent trends include the usage of deep learning [25, 57] and specifically of long-short-term memory-enhanced (LSTM) recurrent NNs [61]. This comes, as the first offer the ability to learn feature representations in an unsupervised manner (as outlined above) which seems attractive in a field where the optimal feature representation is highly disputed. The latter, i.e., the LSTM paradigm, has the charm of providing the additional learning of the optimal temporal context. Again, a valuable asset, as the optimal temporal context for emotion recognition from speech and the best 'unit of analysis' such as fixed length frames, voiced or unvoiced sounds, syllables, words, sequences of words or whole speaker turns is also highly disputed [44, 49].

Integration and Embedding

In this section, we will deal with practical aspects of 'putting SER in there' (i.e., embedding an emotion recogniser in a software system) where 'there' can be a computer game, a (software) application or in principal various kinds of technical, computer, and software systems.

Fusion

Obviously, it will be of interest to exploit both – the spoken content and the way it is acoustically expressed – rather than only any individual aspect of these two. In fact, they provide different strengths when considering emotion in the dimensional model: The words usually better reflect valence aspects, the prosody and acoustics arousal.

The integration of these two cues can be executed either by combining the features before the decision ('early fusion') or after the individual decisions for content and prosody/acoustics ('late fusion'), e.g., by a weighted sum of scores or another classifier/regressor or similar. Further approaches have been investigated such as fusion by prediction, but remained a fringe phenomenon in this domain up to this point.

Available Tools

A number of tools exist that are free for research purposes and often are even open-source.

For modelling emotion and affect via prosody and acoustic features, the openS-MILE [15] package provides a ready to use solution of voice activity detection

and speech signal segmentation, feature extraction, and classification/regression including pre-trained models for an instant setup of an online emotion recogniser. However, it is mostly focused on acoustic/prosodic analysis. A similar tool is described in [58].

For modelling via spoken content analysis, one needs to 'put together' one of the manifold automatic speech recognition engines available these dates such as HTK, Julius, Sphinx, etc. and then plug it into openSMILE or Weka [18], etc., for feature computation and machine learning.

Obviously, further tools exist – in particular for labelling of emotional speech data, e.g., [7] – that could not be mentioned here due to space restrictions.

Data and Benchmarks

Luckily, a richer number of affective speech and language resources is available these days, albeit only few are 'sufficiently' large covering different age groups, and realistic spontaneous emotion. For many languages and cultures, emotionally labelled speech data is unfortunately still missing. Here, only those data that have been featured in a research competition will be named as examples, as they provide a well-defined test-bed and benchmark results are available. These are the FAU Aibo Emotion Corpus (AEC) as was featured in the first comparative public emotion challenge ever – the Interspeech 2009 Emotion Challenge [50]. AEC contains children speech in two or five emotion categories in realistic settings – the best results reached in the Challenge were 71.2 % (two classes) and 44.0 % (five classes) by fusion of the best participants' engines. In 2013, emotion was revisited by the series of challenges in the computational paralinguistic domain by the GEMEP corpus [51]. While it contains enacted data in an artificial syllable 'language', this time 12 classes were targeted with the best result reaching 46.1 %. Another series is the Audio/Visual Emotion Challenge (AVEC) series that has in its so far five editions seen three different databases which are available for research and testing: the AVEC 2011/2012 corpus was based on the SEMAINE database; AVEC 2013/2014's featured AVDLC depression data, and the RECOLA database [37] is featured in AVEC 2015. These are usually larger by an order of magnitude and all labelled in time and value-continuous arousal and valence, and partially further dimensions or classes. A number of further related challenges includes, e.g., [12].

Distribution

The recognition of emotion from speech can usually be done (1) in real-time, and (2) at 'reasonable' computational effort thus (3) allowing for processing on the same platform that is used for a computer game or similar realisation. However, distributing the recognition between a end-user based client and a server bears a

huge advantage: if the audio, the features, or compressed version of the features are collected centrally, learning models can be improved for all players that are making use of such a distributed service. In order to increase the level of privacy of users in this case, and reduce the bandwidth needed for transmission to the server, sub-band feature vector quantisation can be used. In fact, this bears similarity with the generation of the 'audio words' as were described above: Rather than a large feature vector of several thousand features, only one reference ID number of a spatially close reference feature vector is submitted to the server per feature-sub-band and unit of analysis in time [65].

Confidence Measures

The provision of confidence measures gives extra information to a game or application alongside the recognition result on its certainty. This seems needed looking at the current benchmarks in the field as were outlined above. Further to that, confidence measures are an important factor of the self-training as will be described next, as it helps a machine to decide whether it can label data by itself or needs human help prior to using such data to (re-)train itself. Few works have investigated calculation of meaningful confidence measures in this field up to this point. Most notably, these include some aspects tailored to the characteristics of emotion recognition, as they exploit the uncertainty of labels and imbalance of observations per class both typical for the field. The first approach tries to predict the degree of human agreement such as n raters out of N on a data point independent of the emotion assumed as an indication on how reliable the assumption is likely to be. This can be done, as usually several raters (usually some three given expert labellers up to around several tenths, e.g., when crowd sourcing the information) label a data point in this field, and the percentage of agreeing raters can be used as a learning target either instead of the emotion or – even better – alongside the emotion in a multitarget learning approach [16]. The second approach is based on engines trained on additional data with the learning target whether the actual emotion recognition engine is correct or incorrect or – more fancily – correctly positive, correctly negative, incorrectly positive or incorrectly negative [10]. In fact, such confidence measures could be shown to provide reasonable indication on the reliability of a recognised emotion [10] thus providing a genuine extra value that can be exploited, e.g., by a game engine.

Adaptation and Self-Training

In this section, methods are presented on how to exploit existing labelled data similar to the labelled or unlabelled target data or unlabelled data in utmost efficient manners.

In the first case, i.e., the existence of similar data such as adult emotional speech when in need of a child emotion recogniser, e.g., for a child computer game, transfer learning comes in handy. Transfer learning per se is a rather loosely formulated approach in the sense that many different variants and starting points exist. As an example, in [11] a sparse autoencoder is trained in both – the source and the target domain, thus reaching compact representations of both domains. Then, a neural network is trained to learn to map across domains, thus allowing according transfer. In comparison with other standard transfer learning approaches, this has shown the best results in the study for SER. In such a way, a quick adaptation to a new individual, e.g., a new player can be reached.

Further to that, games usually are characterised by repeated interaction with a player. This fact can be exploited to give a game the chance to better 'learn' the user. Ideally, this is done without being noticed or requiring help from the user by weakly supervised learning approaches.

Semi-supervised learning does not require any human help – a system labels the data for its training by itself once it 'hears' (or 'reads') new data and is sufficiently confident that it 'does it right'. This can best be done in co-training exploiting different 'views' on the data such as acoustic and linguistic feature information. This approach has been shown to actually improve recognition performance in the recognition of emotion from speech [28, 67], or text [8] thus enabling systems to also make use of unlabelled data in addition to labelled data.

Unfortunately, it usually requires large amounts of unlabelled data to obtain similar gains in performance as one would see when using (human) labelled data. This can be eased by adding as a promising aid *active learning*. The idea of active learning is to ask a human for help whenever new data appears to be 'promising'. The art here is to quantise and identify what is 'promising'. Approaches that have been shown to work well for emotion recognition are based, e.g., on the likely sparseness of a new unlabelled instance in terms of the expected emotion class. This means that likely to be 'neutral' samples are considered less worthy labelling as there are usually plenty of these available. However, should the system believe a new speech sample is emotional or even potentially from a hardly or never (requiring some additional novelty detection) seen before emotion, it will (only then) ask the player or user about this state, thus considerably reducing the amount of required human labels. Further approaches to measure the interest in a data instance include the expected change in model parameters, i.e., if a new speech sample is not likely to change the parameters of the trained classifier or regressor, it is not important to inquire to which emotion it belongs. Overall, active learning could be shown to reduce the labelling effort by up to 95 % at the same or even higher recognition accuracy of accordingly trained models in the study [66].

Most efficiently, *cooperative learning* combines the strengths and increases the efficiency of active and semi-supervised learning in the sense that a computer system first decides if it can label the data itself, and only if not, evaluates if it is worth to ask for human help [64].

Encoding and Standards

A number of standards exist to encode the information on emotion to be used by a game or application. To name but a few, the Humaine EARL and the W3C Emotion Markup Language recommendation (or EmotionML [42] for short) have been designed in particular for encoding affective information. A number of further standards do, however, provide tags for emotion as well, and may likely be used in a video game or similar context, anyhow, such as W3C EMMA [1].

Further, in particular for the speech modality, a standard has been developed to encode the feature information used [4]. This allows to exchange collected feature/label pairs even if different systems or games use different types of feature extraction. In addition, standards exist for the encoding of individual feature types, e.g., [55].

Finally, standardised feature sets are offered and used in this field by the openSMILE toolkit (cf. above).

Summary and White Spots

Towards the end of this chapter, let us quickly summarise and identify white spots left for future research efforts.

Summary

In this chapter, we took a look 'under the hood' of a modern speech emotion recognition engine. Ideally, it exploits both, acoustic and linguistic cues and combines these efficiently. The features can either be extracted based on known 'expert' feature types or learnt unsupervised from data, e.g., in deep learning. Then, either a rule-based 'zero-resource' decision takes place, or a machine learning algorithm is trained based on labelled speech data. The recognition result for a previously unseen new speech instance will then usually be one or several emotion labels or one or several continuous values of emotion primitives such as arousal or valence. What's more, an engine should ideally provide confidence measures alongside to give a feeling for the reliability of its prediction. It can, based on this confidence measure, also decide to retrain itself after having heard new data points and having been able to predict them likely correctly. If it is unsure about the emotion of a new sample but feels it would benefit from knowing it, it can ask a user for help in labelling in a very targeted and efficient way. If it has 'similar' labelled data to the target data to exploit or start with, transfer learning allows to adapt the data to the domain. For practical implementation, encoding and feature standards can be used, and a number of 'out of the box' ready tools exist which are free for

research purposes. There have repeatedly been evaluation campaigns showing that recognition of emotion from speech 'works', but leaves head room for improvement at this moment in time.

White Spots

The field is still somewhat young and not all problems could be touched upon sufficiently or at all, yet. To conclude this chapter, these white spots in the literature shall be quickly named without lengthy explanation: When dealing with textual cues, *multilinguality* is an obvious problem [2, 36]. While one can benefit from the progress machine translation has made, there is still need for more experience with this challenge. Obviously, multilinguality also has an impact on acoustic features, e.g., when considering the differences between tonal or non-tonal languages such as Chinese in comparison to English or German. Similarly, *cross-cultural* aspects [14, 39, 54] need additional investigation and in particular solutions to identify the culture automatically and choose or adapt models accordingly, e.g., by suited means of transfer learning. Further, 'faking', *regulation*, and *masking* of emotions has hardly been investigated in terms of automatic recognition. Similarly, '*atypicality*' is not sufficiently explored up to this moment. This deficit includes the lack of providing an engine prepared for less typical emotion portrayal as, e.g., from individuals with ADHD, Autism Spectrum Condition, vocal disfluencies, and alike or simply of less investigated age groups such as the elderly or (very) young individuals. Obviously, a number of *ethical implications* come with such data, and in general when having a machine analysing human emotion, giving (potentially flawed) feedback on it, and potentially distributing processing or storing information in a centralised fashion [3, 13, 56]. Then, *user studies* and experiences with *real systems 'in the wild'* are still very sparse and among most urgent issues to tackle – often a careful system or game design or response pattern can cover up very elegantly for potential misrecognitions, or – the other way round – an almost perfectly functioning emotion recognition not exploited in the right way can 'ruin it all' easily. Finally, better modelling of *co-influencing states and traits* of humans of analysis is needed: All speaker states and traits impact on the same speech production mechanism and the choice of our verbal behaviour: Whether we have a cold, are sleepy, or intoxicated – emotion should be recognised reliably independent of these factors and our personality profile. A promising route to reaching this ability seems parallel modelling of the wide range of state and traits, e.g., by a neural network with multiple output nodes as targets for various speaker states and traits rather than just for emotion, yet, allowing for missing labels during training, as not all data are likely to be labelled in such a wide range of factors.

With the advent of solutions to these problems as well as in general, one can expect to see gaming and general intelligent technology soon to be recognising our emotion when having us talk (or type) to them. May the best use be made of this new technical ability considering highest ethical standards at all time.

References

 1. Baggia P, Burnett DC, Carter J, Dahl DA, McCobb G, Raggett D (2007) Emma: extensible multimodal annotation markup language. W3C working draft
 2. Banea C, Mihalcea R, Wiebe J (2011) Multilingual sentiment and subjectivity analysis. Multiling Nat Lang Process 6:1–19
 3. Batliner A, Schuller B (2014) More than fifty years of speech processing – the rise of computational paralinguistics and ethical demands.   In: Proceedings ETHICOMP 2014. CERNA, Paris, 11p
 4. Batliner A, Steidl S, Schuller B, Seppi D, Vogt T, Wagner J, Devillers L, Vidrascu L, Aharonson V, Kessous L, Amir N (2011) Whodunnit – searching for the most important feature types signalling emotion-related user states in speech.   Comput Speech Lang Spec Issue Affect Speech Real-life Interact 25(1):4–28
 5. Becker C, Nakasone A, Prendinger H, Ishizuka M, Wachsmuth I (2005) Physiologically interactive gaming with the 3D agent max.   In: Proceedings international workshop on conversational informatics in conjunction with JSAI-05, Kitakyushu, pp 37–42
 6. Brückner R, Schuller B (2014) Social signal classification using deep BLSTM recurrent neural networks. In: Proceedings 39th IEEE international conference on acoustics, speech, and signal processing, ICASSP 2014, Florence. IEEE, pp 4856–4860
 7. Cowie R, Douglas-Cowie E, Savvidou S, McMahon E, Sawey M, Schröder M (2000) Feeltrace: an instrument for recording perceived emotion in real time. In: Proceedings ISCA workshop on speech and emotion, Newcastle, pp 19–24
 8. Davidov D, Tsur O, Rappoport A (2010) Semi-supervised recognition of sarcastic sentences in Twitter and Amazon. In: Proceedings CoNNL, Uppsala, pp 107–116
 9. Dellaert F, Polzin T, Waibel A (1996) Recognizing emotion in speech. In: Proceedings ICSLP, Philadelphia, pp 1970–1973
10. Deng J, Schuller B (2012) Confidence measures in speech emotion recognition based on semi-supervised learning. In: Proceedings of INTERSPEECH. ISCA, Portland
11. Deng J, Zhang Z, Schuller B (2014) Linked source and target domain subspace feature transfer learning – exemplified by speech emotion recognition.   In: Proceedings 22nd international conference on pattern recognition (ICPR 2014). IAPR, Stockholm, pp 761–766
12. Dhall A, Goecke R, Joshi J, Wagner M, Gedeon T (eds) (2013) Proceedings emotion recognition in the wild challenge and workshop. ACM, Sydney
13. Döring S, Goldie P, McGuinness S (2011) Principalism: a method for the ethics of emotion-oriented machines. In: Petta P, Pelachaud C, Cowie R (eds) Emotion-oriented systems: the HUMAINE handbook, cognitive technologies. Springer, Berlin/Heidelberg, pp 713–724
14. Elfenbein HA, Mandal MK, Ambady N, Harizuka S, Kumar S (2002) On the universality and cultural specificity of emotion recognition: a meta-analysis. Psychol Bull 128(2):236–242
15. Eyben F, Weninger F, Groß F, Schuller B (2013) Recent developments in openSMILE, the Munich open-source multimedia feature extractor.   In: Proceedings of the 21st ACM international conference on multimedia, MM 2013. ACM, Barcelona, pp 835–838
16. Eyben F, Wöllmer M, Schuller B (2012) A multi-task approach to continuous five-dimensional affect sensing in natural speech. ACM Trans Interact Intell Syst Spec Issue Affect Interact Nat Environ 2(1):29
17. Gao Y, Bianchi-Berthouze N, Meng H (2012) What does touch tell us about emotions in touchscreen-based gameplay? ACM Trans Comput-Hum Interact 19(4/31):1–30
18. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I (2009) The WEKA data mining software: an update. SIGKDD Explor 11:10–18

19. Holmgard C, Yannakakis G, Karstoft KI, Andersen H (2013) Stress detection for PTSD via the StartleMart game. In: Proceedings of 2013 humaine association conference on affective computing and intelligent interaction (ACII). IEEE, Memphis, pp 523–528
20. Hudlicka E (2009) Affective game engines: motivation and requirements. In: Proceedings of the 4th international conference on foundations of digital games. ACM, New York, 9p
21. Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. In: Nédellec C, Rouveirol C (eds) Proceedings of ECML-98, 10th European conference on machine learning. Springer, Heidelberg/Chemnitz, pp 137–142
22. Johnstone T (1996) Emotional speech elicited using computer games. In: Proceedings ICSLP, Philadelphia, 4p
23. Johnstone T, van Reekum CM, Hird K, Kirsner K, Scherer KR (2005) Affective speech elicited with a computer game. Emotion 5:513–518
24. Kim J, Bee N, Wagner J, André E (2004) Emote to win: affective interactions with a computer game agent. In: GI Jahrestagung, Ulm, vol 1, pp 159–164
25. Kim Y, Lee H, Mower-Provost E (2013) Deep learning for robust feature generation in audiovisual emotion recognition. In: Proceedings the 2nd CHiME workshop on machine listening in multisource environments held in conjunction with ICASSP 2013, Vancouver. IEEE, pp 86–90
26. Liscombe J, Hirschberg J, Venditti JJ (2005) Detecting certainness in spoken tutorial dialogues. In: Proceedings INTERSPEECH. ISCA, Lisbon, pp 1837–1840
27. Litman D, Forbes K (2003) Recognizing emotions from student speech in tutoring dialogues. In: Proceedings ASRU, Virgin Island. IEEE, pp 25–30
28. Mahdhaoui A, Chetouani M (2009) A new approach for motherese detection using a semi-supervised algorithm. In: Machine learning for signal processing XIX – Proceedings of the 2009 IEEE signal processing society workshop, MLSP 2009, Grenoble. IEEE, pp 1–6
29. Martyn C, Sutherland JJ (2005) Creating an emotionally reactive computer game responding to affective cues in speech. In: Proceedings HCI, Las Vegas, vol 2, pp 1–2
30. Metze F, Batliner A, Eyben F, Polzehl T, Schuller B, Steidl S (2010) Emotion recognition using imperfect speech recognition. In: Proceedings INTERSPEECH. ISCA, Makuhari, pp 478–481
31. Missen M, Boughanem M (2009) Using WordNet's semantic relations for opinion detection in blogs. In: Advances in information retrieval. Lecture notes in computer science, vol 5478/2009. Springer, Berlin, pp 729–733
32. Mower E, Mataric MJ, Narayanan SS (2011) A framework for automatic human emotion classification using emotion profiles. IEEE Trans Audio Speech Lang Process 19:1057–1070
33. Pachet F, Roy P (2009) Analytical features: a knowledge-based approach to audio feature generation. EURASIP J Audio Speech Music Process 2009:1–23
34. Park S, Sim H, Lee W (2014) Dynamic game difficulty control by using EEG-based emotion recognition. Int J Control Autom 7:267–272
35. Ploog BO, Banerjee S, Brooks PJ (2009) Attention to prosody (intonation) and content in children with autism and in typical children using spoken sentences in a computer game. Res Autism Spectr Disord 3:743–758
36. Polzehl T, Schmitt A, Metze F (2010) Approaching multi-lingual emotion recognition from speech – on language dependency of acoustic/prosodic features for anger detection. In: Proceedings speech prosody, Chicago. ISCA
37. Ringeval F, Eyben F, Kroupi E, Yuce A, Thiran JP, Ebrahimi T, Lalanne D, Schuller B (2015) Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data. Pattern Recognit Lett 66:10
38. Rudra T, Kavakli M, Tien D (2007) Emotion detection from female speech in computer games. In: Proceedings of TENCON 2007 – 2007 IEEE region 10 conference, Taipei. IEEE, pp 712–716
39. Sauter DA, Eisner F, Ekman P, Scott SK (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. Proc Natl Acad Sci USA 107(6):2408–2412
40. Scherer KR (2003) Vocal communication of emotion: a review of research paradigms. Speech Commun 40:227–256

41. Scherer S, Hofmann H, Lampmann M, Pfeil M, Rhinow S, Schwenker F, Palm G (2008) Emotion recognition from speech: stress experiment. In: Proceedings of the international conference on language resources and evaluation, LREC 2008, Marrakech. ELRA, 6p
42. Schröder M, Devillers L, Karpouzis K, Martin JC, Pelachaud C, Peter C, Pirker H, Schuller B, Tao J, Wilson I (2007) What should a generic emotion markup language be able to represent? In: Paiva A, Prada R, Picard RW (eds) Proceedings of ACII. Springer, Berlin/Heidelberg, pp 440–451
43. Schuller B (2012) The computational paralinguistics challenge. IEEE Signal Process Mag 29(4):97–101
44. Schuller B, Batliner A (2013) Computational paralinguistics: emotion, affect and personality in speech and language processing. Wiley, New York
45. Schuller B, Devillers L (2010) Incremental acoustic valence recognition: an inter-corpus perspective on features, matching, and performance in a gating paradigm. In: Proceedings INTERSPEECH, Makuhari. ISCA, pp 2794–2797
46. Schuller B, Dunwell I, Weninger F, Paletta L (2013) Serious gaming for behavior change – the state of play. IEEE Pervasive Comput Mag 12(3):48–55
47. Schuller B, Knaup T (2011) Learning and knowledge-based sentiment analysis in movie review key excerpts. In: Esposito A, Esposito AM, Martone R, Müller V, Scarpetta G (eds) Toward autonomous, adaptive, and context-aware multimodal interfaces: theoretical and practical issues: third COST 2102 international training school. Lecture notes on computer science (LNCS), vol 6456/2010, 1st edn. Springer, Heidelberg, pp 448–472
48. Schuller B, Marchi E, Baron-Cohen S, Lassalle A, O'Reilly H, Pigat D, Robinson P, Davies I, Baltrusaitis T, Mahmoud M, Golan O, Friedenson S, Tal S, Newman S, Meir N, Shillo R, Camurri A, Piana S, Staglianò A, Bölte S, Lundqvist D, Berggren S, Baranger A, Sullings N, Sezgin M, Alyuz N, Rynkiewicz A, Ptaszek K, Ligmann K (2015) Recent developments and results of ASC-inclusion: an integrated internet-based environment for social inclusion of children with autism spectrum conditions. In: Proceedings of the of the 3rd international workshop on intelligent digital games for empowerment and inclusion (IDGEI 2015) as part of the 20th ACM international conference on intelligent user interfaces, IUI 2015, Atlanta. ACM, 9p
49. Schuller B, Rigoll G (2006) Timing levels in segment-based speech emotion recognition. In: Proceedings of INTERSPEECH, Pittsburgh. ISCA, pp 1818–1821
50. Schuller B, Steidl S, Batliner A (2009) The interspeech 2009 emotion challenge. In: Proceedings of INTERSPEECH, Brighton. ISCA, pp 312–315
51. Schuller B, Steidl S, Batliner A, Vinciarelli A, Scherer K, Ringeval F, Chetouani M, Weninger F, Eyben F, Marchi E, Mortillaro M, Salamin H, Polychroniou A, Valente F, Kim S (2013) The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. In: Proceedings of INTERSPEECH, Lyon. ISCA, pp 148–152
52. Schuller B, Zhang Z, Weninger F, Burkhardt F (2012) Synthesized speech for model training in cross-corpus recognition of human emotion. Int J Speech Technol Spec Issue New Improv Adv Speak Recognit Technol 15(3):313–323
53. Shahid S, Krahmer E, Swerts M (2007) Audiovisual emotional speech of game playing children: effects of age and culture. In: Proceedings of INTERSPEECH, Antwerp, pp 2681–2684
54. Shaver PR, Wu S, Schwartz JC (1992) Cross-cultural similarities and differences in emotion and its representation: a prototype approach. Review of Personality and Social Psychology. Vol. XIII: Emotion, 175–212
55. Silverman K, Beckman M, Pitrelli J, Ostendorf M, Wightman C, Price P, Pierrehumbert J, Hirschberg J (1992) ToBI: a standard for labeling English prosody. In: Proceedings of ICSLP, Banff, pp 867–870
56. Sneddon I, Goldie P, Petta P (2011) Ethics in emotion-oriented systems: the challenges for an ethics committee. In: Petta P, Pelachaud C, Cowie R (eds) Emotion-oriented systems: the HUMAINE handbook, cognitive technologies. Springer, Berlin/Heidelberg, pp 753–768

57. Stuhlsatz A, Meyer C, Eyben F, Zielke T, Meier G, Schuller B (2011) Deep neural networks for acoustic emotion recognition: raising the benchmarks. In: Proceedings 36th IEEE international conference on acoustics, speech, and signal processing, ICASSP 2011, Prague. IEEE, pp 5688–5691

58. Vogt T, André E, Bee N (2008) Emovoice – a framework for online recognition of emotions from voice. In: Proceedings IEEE PIT, Kloster Irsee. Lecture notes in computer science, vol 5078. Springer, pp 188–199

59. Weninger F, Schuller B, Batliner A, Steidl S, Seppi D (2011) Recognition of non-prototypical emotions in reverberated and noisy speech by non-negative matrix factorization. EURASIP J Adv Signal Process Spec Issue Emot Ment State Recognit Speech 2011:Article ID 838790

60. Weninger FJ, Watanabe S, Tachioka Y, Schuller B (2014) Deep recurrent de-noising auto-encoder and blind de-reverberation for reverberated speech recognition. In: Proceedings 39th IEEE international conference on acoustics, speech, and signal processing, ICASSP 2014, Florence. IEEE, pp 4656–4660

61. Wöllmer M, Schuller B, Eyben F, Rigoll G (2010) Combining long short-term memory and dynamic Bayesian networks for incremental emotion-sensitive artificial listening. IEEE J Select Top Signal Process Spec Issue Speech Process Nat Interact Intell Environ 4(5):867–881

62. Yildirim S, Lee C, Lee S, Potamianos A, Narayanan S (2005) Detecting politeness and Frustration state of a child in a conversational computer game. In: Proceedings of INTERSPEECH, Lisbon. ISCA, pp 2209–2212

63. Yildirim S, Narayanan S, Potamianos A (2011) Detecting emotional state of a child in a conversational computer game. Comput Speech Lang 25:29–44

64. Zhang Z, Coutinho E, Deng J, Schuller B (2015) Cooperative learning and its application to emotion recognition from speech. IEEE/ACM Trans Audio Speech Lang Process 23(1):115–126

65. Zhang Z, Coutinho E, Deng J, Schuller B (2014) Distributing recognition in computational paralinguistics. IEEE Trans Affect Comput 5(4):406–417

66. Zhang Z, Deng J, Marchi E, Schuller B (2013) Active learning by label uncertainty for acoustic emotion recognition. In: Proceedings of INTERSPEECH, Lyon. ISCA, pp 2841–2845

67. Zhang Z, Deng J, Schuller B (2013) Co-training succeeds in computational paralinguistics. In: Proceedings 38th IEEE international conference on acoustics, speech, and signal processing, ICASSP 2013, Vancouver. IEEE, pp 8505–8509

Chapter 6
Comparing Two Commercial Brain Computer Interfaces for Serious Games and Virtual Environments

Szymon Fiałek and Fotis Liarokapis

Abstract Brain-Computer Interface (BCI) technology is still under development, however the recent advances allowed to move BCI from research laboratories to people's living rooms. One of the promising areas of BCI applications is in computer games and virtual environments. In this chapter, initially an overview of the state of the art of BCI applications in computer games is presented. Next, a user study of two inexpensive commercially available devices used in different games is presented. The results indicate that multi-channel BCI systems are better suited for controlling an avatar in 3D environments in an active manner, while BCI systems with one channel is well suited for use with games utilising neuro-feedback. Finally, the findings demonstrate the importance of matching appropriate BCI devices with the appropriate game.

What Are Brain-Computer Interfaces?

A brain computer interface (BCI), also known as brain-machine interface (BMI) is a system that allows for direct communication between a human and a machine without using traditional channels of interaction, e.g. the muscles of the arm and hand and a keyboard, and instead relies on brain signals directly [30]. This makes BCI technology especially attractive for people with severe motor disabilities such as multiple sclerosis (MLS) or locked in syndrome [3]. In extreme cases such interface is the only way by which a person can communicate with the external world, which can greatly improve their quality of life. The idea of BCI was initially unattractive to science, the idea of deciphering human thoughts seemed weird and remote. BCI systems were limited to laboratory and clinical use; however the recent developments in machine learning technology and increase in computational power of personal computers made BCI accessible not only to researchers and clinicians,

S. Fiałek (✉) • F. Liarokapis
HCI Lab, Faculty of Informatics, Masaryk University, Brno, Czech Republic
e-mail: sfialek@mail.muni.cz; liarokap@fi.muni.cz

but also for everyday users. BCI systems can be categorised into invasive and non-invasive. In case of invasive approach the signal is acquired using electrodes placed inside the scalp. In non-invasive BCI the signal is acquired outside of the scalp. In this chapter, the results of investigation into the use of two commercially available BCI devices for gaming and virtual environments are presented. In section "Neuroimaging Techniques for BCI Systems", BCI systems based on different neuroimaging techniques are described. Section "Electroencephalography-Based BCI (EEG-Based BCI)", presents the theoretical bases for different EEG-based BCIs. Stages of BCI systems are illustrated in section "Stages of a BCI System". Section "BCI Paradigms" demonstrates presents different BCI paradigms used in BCI games, followed by different areas of BCI applications in section "BCI and Serious Games". Finally, the description of performed experiments is presented in sections "Investigating Commercial BCI Systems for Serious Games and Virtual Environments" and "Conclusions and Limitations" presents conclusions.

Neuroimaging Techniques for BCI Systems

Different indicators of brain activity as well as different neuroimaging techniques can be used for signal acquisition and these are described below.

There are two types of brain activity that can be used for this purpose: metabolic and electromagnetic. Only the latter type of systems is currently available for day-to-day BCI devices. For each of the brain activity modalities there are two different neuroimaging techniques that can be used. For BCIs based on metabolic brain activity we can distinguish two types of systems functional Magnetic Resonance Imaging (fMRI) based BCIs and functional Near Infrared Spectroscopy (fNIRS) based BCIs. For BCIs based on electromagnetic brain activity we can distinguish two types of systems magnetoencephalography (MEG) based BCIs and electroencephalography (EEG) based BCIs. These systems are discussed below.

For the first category, both rely on very different method of signal acquisition, and are consequently sensitive to very different variations in the signal. fMRI is primarily used in research, because it is not practical for day-to-day activities, whereas fNIRS, which may be more accessible to a wider audience, is still under development.

MEG is a non-invasive imaging technique that registers changes in magnetic field associated with the electrical activity of the brain. MEG and EEG record signal associated with the same neurophysiological processes. However, the magnetic field is less prone to distortions introduced by the skull and the scalp than the electric field, therefore the quality of signal provided by the MEG is better than in case of EEG. The main limitations of MEG-based BCI are similar to the ones of fMRI-based systems. These include high cost and large equipment that cannot be used outside of a laboratory.

Electroencephalography-Based BCI (EEG-Based BCI)

EEG is the oldest and most widely used neuroimaging technique. Since its discovery in 1929 [2], EEG has been used by scientists to answer question about the functioning of the human brain as well as by clinicians as a diagnostic tool. BCI systems also allow using EEG as neurofeedback in neurorehabilitation. One of the reasons for the popularity of EEG based systems is the relative low cost, portability and low complexity of these devices.

EEG signal collected on the surface of the skull is a result of the activity of neurons. Neuron is a call in the nervous system that processes and transmits information using electrical and chemical signals. A typical neuron possesses a cell body (soma), dendrites, and an axon. The cell body can give rise to many dendrites. Dendrites are thin structures that span for hundreds of microns and form a dendritic tree. The cell body can give rise to multiple dendrites, but can produce only one axon, however axons can branch out hundreds of times before they terminate. A human axon can extend for up to one meter. Dendrites and axons connect neurons forming neural networks. Information from one neuron to another is passed through synapses. In most cases a synapse connect an axon to a dendrite, however there are exceptions (e.g. axon can connect directly to the body of the neuron).

Neurons communicate through action potentials, i.e. electrical discharge produced by the soma of the cell. Action potential travels along the axon and when the action potential arrives at the synapses neurotransmitter is released. Neurotransmitter triggers change in potential of the membrane of the receiving cell (flow of ions through the cell membrane) and if this potential reaches a threshold, new action potential is triggered and the information is transmitted to another neuron.

The signal measured using EEG equipment is thought to be generated mostly by the pyramidal neurons located in the cerebral cortex [14]. Pyramidal neurons have large soma of a shape that resembles a pyramid and a large dendrite extending from the apex of the soma and is directed perpendicular to the surface of the cortex. Activation of an excitatory synapse creates excitatory post-synaptic potential, i.e. inflow of positively charged ions form the extracellular space to body of the neuron. As a result, the extracellular region of the synapse becomes negatively charged and in turn regions distant from the synapse become positively charged and cause a change of potential (extracellular current) to flow towards the region of the synapse. The spatio-temporal summation of these extracellular currents at hundreds of thousands of neurons with parallelly oriented dendrites creates the change of potential that is detectable on the surface of the scalp. If a large number of excitatory synapses are activated close to the surface of the cortex, the resulting potential, detectable on the surface of the scalp, is negative. If synapses of the same type are activated closer to the body of the pyramidal neurons, deeper in the cortex the resulting potential is positive. Reverse relation is observed for inhibitory synapses. Activation of a large number of inhibitory synapses close to the surface of the brain produces positive potential and activation of inhibitory synapses in the deeper layers

of the cortex results in negative potentials recordable on the surface of the scalp. It is therefore possible to infer the type of synapses activated form the polarity of the signal acquired on the surface of the scalp.

EEG is the recording of electrical potentials along the scalp. EEG recordings are usually performed using small metal electrodes placed on the scalp in standardised positions. The number of electrodes can vary depending on the type of device used. To improve the conductivity between scalp and electrodes conductive gel or saltwater are used. The electric potentials recorded by EEG result for the neural activation within the brain. EEG is the most widespread neuroimaging technique and the most widely used modality in BCI. The popularity of EEG stands form the fact that electrical signal can be easily and cheaply recorded through electrodes placed on the scalp [1]. However, electric current has to cross the scalp, skull and other tissues surrounding the brain which significantly distorts the acquired signal. EEG signal is also distorted by the electrical noise in the environment and electric current produced by muscles.

Researches using EEG identified distinctive patterns in the EEG signal. These patterns are related to specific cognitive activities performed. Although, the exact meaning of most of these patterns is still unknown, some of them have been thoroughly studied and are used in BCI systems. The three EEG signal features that are most often used in BCI research are P300, sensorimotor rhythms and steady state visually evoked potentials [25].

Stages of a BCI System

BCI can be considered as artificial intelligence system that employs machine learning. Such a system consists of hardware and software components with the aim of recognising patterns in the signals emitted by the brain, and to translate them into practical commands. In a typical BCI system, five consecutive stages can be identified and these are presented below [25].

1. Signal acquisition – various types of signal are captured by a neuroimaging device such as electroencephalography (EEG); a BCI system may be acquiring several kinds of signals at the same time, provided they are synchronised and time-locked to the interaction with the device.
2. Signal pre-processing or signal enhancement – signal is prepared to further processing, including artefact removal (e.g. muscle movement, and noise reduction) are typically performed at this stage.
3. Feature extraction – discriminative features are identified and mapped onto a vector; these may include first order parameters, like amplitude of signal or latency, and second-order parameters that require more processing, like time-frequency parameters extracted from a Fourier transform.
4. Classification – involves the classification of the parameters previously extracted, with the aim of ascribing meaning to them; various techniques from machine

learning can be applied, but this imposes an overhead in time and processing power that is not suitable to all BCI applications, which demands real-time interaction.

5. Control interface – results of classification are translated into commands and send to a connected machine such as a wheelchair or a computer, which provide the user with feedback and close the interactive loop between the user and the device.

These stages create a communication loop between the user and the machine which is the essence of BCI. Different ways of implementing this loop can be proposed and this can result in different control paradigms: active, passive and reactive. Below, we will discuss these paradigms in the context of computer games.

BCI Paradigms

Active BCI

In case of active BCI user modulates brain signal actively. In order to effectively control the computer this signal should be discriminative. Active BCI can be used to directly control an BCI application. Active BCI applications often make use of motor imagery as the control paradigm. Imagination of motor movements result in contra-lateral event related de-synchronization (ERD), and ipsi-latelar event related synchronisation (ERS). When the user imagines a right hand movement, the amplitude of the mu-rhythm in the left sensory-motor area increases while the amplitude of mu-rhythm in the right sensory-motor area decreases. These changes can be observed in mu-rhythm (7.5–12.5 Hz). An active BCI pinball game was presented by [26]. Pinball is a fast-paced game with rich environment that requires very fast and precisely timed reactions. The users were able to achieve good control of the game and rated their experience very highly.

Reactive BCI

In reactive BCI, the brain signal containing information is generated as a response to external stimulation. The user introduces changes in the signal by attending to the stimuli. Reactive BCI paradigms include steady state visual evoked potential (SSVEP) and P300. In SSVEP, the user concentrate on one of flickering stimuli. This results in an increase in activity of the same frequency (or a harmonic of that frequency) observed in visual areas and allows to identify the stimuli to which the user is attending to. Jackson et al. [8] proposed a first person shooter type game in which the user could move around fire the gun by concentrating their eyes on one of four flashing stimuli displayed on the screen.

In P300-based BCI systems, the user concentrates on one of many randomly activated stimuli. The P300 (also called P3) wave is a component of an event related potential (ERP) which is elicited by infrequent auditory, visual or somatosensory stimuli. P300 is a positive peak in EEG signal occurring around 300 msec after onset of the event. This change in EEG signal is normally elicited using oddball paradigm, as a response to a low probability stimulus that appears amongst high probability stimuli. P300 is also augmented when one perceives a stimulus that is regarded to be important; a stimulus one pays attention to. A P300-based BCI game, so called Mind Game was developed by [5]. It is a 3D checkerboard-styled board game with aim of visiting all trees placed on the board. The trees are also used as a target like in typical P300 seller paradigm, the squares underneath the trees are illuminated randomly. The user moves an avatar on the board by gazing their eyes on the target tree.

Passive BCI

The primary objective of passive BCI system is not to provide the user with the ability to intentionally control the device, but to monitor the user's mental state. Using these systems does not require much effort the monitoring of the users mental state happens automatically. In passive BCI, the primary role of the system is not to give control to the user, and it does not require any effort on the user's part; instead, it monitors the user's mental states automatically. The level of desired mental state is quantified and used to facilitate the communication between the user and the system.

One of the most popular computer games has been adapted for BCI input. In Alpha-World of Warcraft [27] the level of alpha activity, recorded over the parietal lobe, was used to control one aspect of the game – the type of the avatar, other aspects of this game were controlled using traditional methods (keyboard and computer mouse). High level of alpha activity over the parietal lobe is believed to be indicative of relaxed alertness which is probably the bast state of mind for playing computer games. On the contrary, it was assumed that a low level of alpha indicated the state of distress. When the level of alpha was high the avatar would assume the form of a druid and a low level of alpha resulted the avatar changing its form to a bear, with sharp teeth and claws. In time of distress, the bear form of avatar was better suited for a fight, while in relaxed times the druid form was more fragile, but able to cast effective spells as well as heal herself.

BCI and Serious Games

This section describes different areas of application domains in which BCI games have been used. These include: research, medicine and commercial BCI games.

BCI Games for Research

EEG technology has been use in clinical and research applications for decades. Adding a BCI and gaming components greatly enhance the research possibilities. The first BCI games designed 1977 was used exactly for these purposes [28]. In the experiment the participants navigated a 2D maze using ERP differences in EEG signal, much like P300 paradigm. However, in this care the researchers analysed the negative change which occurred 100 msec. after the stimulus presentation. The users concentrated their eyes on one of four stimuli (top, down, left, right) and this allowed them to move within the virtual maze. A number of well known games like Packman, Pong and Tetris have been adapted to be used with Berlin brain-computer interface (BBCI) [10]. Other BCI games developed for research purposes include flight simulator [15] and a tight rope walker controlled to the changes in left and right hemisphere activity [21]. Reuderink et al. [19] developed a game based on packman. The game was used to induce frustration and investigate the influence of frustration on the quality if EEG signal. The preliminary results indicate that frustration may have an effect on the quality of EEG signal and may affect the signal features used for classification. Frustration and tiredness often leads to decrease in BCI performance. Research into the effects of embedding BCI in games found the opposite effects. User performance increased when playing BCI game, this is most likely due to the fact that BCI games offer reach environment which increases the entertainment that stimulate and motivates the users.

BCI Games for Medical Applications

Neuro-feedback training is an alternative to medication for treatment of some neurological disorders. The most popular of them is attention deficit hyperactive disorder (ADHD). A number of applications using BCI for this purpose have been developed. Most of them rely on modulation of slow cortical potentials. This methods have been used inn the clinical settings for decades [20]. The availability of inexpensive BCI devices made it possible to use these techniques at home. The main obstacle in popularisation of this type of neuro-feedback is the lack of interest form the potential patients – often children who due to their condition find if difficult to concentrate on performing boring and mundane tasks. The application of computer game technology can makes the training more entertaining and encourage children to engage in it. Strehl et al. [24] developed a game that allowed users to steer a ball to a target using modulation of slow cortical potentials. Another example of a passive BCI device used for medical purposes was developed by [18]. The researchers investigated the use of neuro-feedback for the treatment of ADHD. The participants were divided into two groups. One group used standard neuro-feedback, while the other used neuro-feedback embedded in a computer game. Correct brain activity pattern was rewarded by more responsive controls.

Commercial BCI Games

Large game companies like Microsoft, Sony and Nintendo are yet to move into the world of BCIs. The market is no dominated by, however very active players. Two probably most popular producers of commercially available BCI systems are Emotiv [4] and NeuroSky [17].

The game available commercially can be divided into two categories, those offering the participants the ability to control a an avatar in virtual environment and those using neuro-feedback for the purpose of altering the users mental state and thereby improving user's well-being. The 'control' applications employ active paradigm and usually require training. An example of one of them is StoneHenge game developed by Emotiv [23]. In this game users are requested to rebuilt the Stonehange, by using motor imagery to lift, push, pull and rotate giant stones until they are placed in the desired position. Another game developed by Emotiv is 'Mindala' [16] in which users can train meditation skills by controlling a mandala displayed on the computer screen. This is an example of neuro-feedback, passive application. In order for BCI to become a popular amongst computer game enthusiasts relatively inexpensive BCI devices have to provide good gamer experience. In the next section we evaluate two inexpensive and very popular currently available device.

Investigating Commercial BCI Systems for Serious Games and Virtual Environments

This section presents the evaluation of BCI systems using two commercially available devices: NeuroSky and Emotiv. Firstly, we present the evaluation of two BCI systems utilising Emotiv devive section "Comparison of Serious Games Using Emotiv". Secondly, the evaluation of two BCI systems based on NeuroSky is presented section "Comparison of Serious Games Using NeuroSky". Finally, a presentation of these devices is presented (Fig. 6.1).

The Emotiv headset [4] is a neuro-signal acquisition and processing wireless neuro-headset with 14 wet sensors (and 2 reference sensors) which is capable of detecting EEG signal as well user's facial expressions. The neural signal is analysed and a number of affective measures are provided, these include 'engagement', 'frustration', 'meditation' and 'excitement'. This type of measures can be utilized in a passive BCI design. Moreover, a built-in classifier can be trained using different mental activity (e.g. motor imagination) and these can be assigned to operations in the virtual environment, such as 'push', 'pull', 'rotate', and 'lift'. These parameters can be easily used for an active BCI application.

The NeuroSky MindSet [17] device is a wireless headset with speakers and a microphone and one eeg sensor place on the forehead. Most of the signal collected from by this sensor corresponds to the frontal lobe, which limits the types of mental activity that can be used for controlling BCI. NeuroSky provide two measures

Fig. 6.1 BCI devices. On the *left* – Emotiv Epoch; image available at http://emotiv.com, On the *right* – Neurosky Mindset; image available at http://neurosky.com. Epoch is a trademark of Emotiv, Mindset is a trademark of Neurosky (Images used with permission)
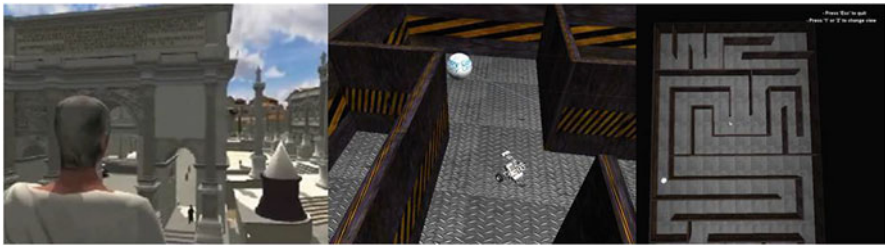
Fig. 6.2 BCI games used with Emotiv device: from the *left* – Roma Nova, LEGO NXT Robot, LEGO NXT Robot (view of the labyrinth from the top) (Images used with permission from the Serious Games Institute, Coventry University)

precomputed by a 'black-boxed' NeuroSky algorithm. These are 'attention' and 'meditation'. The measure of attention is a modulation of the frequency band which is triggered by the intensity of the user's level of mental 'focus' when user focuses. It increases when the user concentrates on a single thought or an external object and decreases when the user is distracted. Both of the measures are well suited for passive BCI applications.

Comparison of Serious Games Using Emotiv

In this section we present the evaluation of two BCI systems which were develop with Emotiv device; the BrainMaze Game and Roma Nova (see Fig. 6.2).

The 'BrainMaze' Game

BrainMaze was designed for the user to navigate a 3D version of the LEGO NXT Robot [29] inside a maze with main goal to find the different waypoints that will

lead to the finish line. If the robot hits a wall, the position resets and starts again from the beginning. Users have to be precise and cautious in order to find the way to the end. The walls are relatively narrow, requiring precise control and avoidance of sudden movements that could cause a position reset. The game session follows the first training session, during which the user trained in the Control Panel, so that the user is familiar with basic brain control.

Roma Nova

Roma Nova is built upon Rome Reborn [6] one of the most realistic 3D representations of Ancient Rome currently in existence. This 3D representation provides a high fidelity 3D digital model which can be explored in real-time. Rome Reborn includes hundreds of buildings, thirty two of which are highly detailed monuments reconstructed on the basis of reliable archaeological evidence. The interactive game is a serious game that aims at teaching history to children (11–14 years old). The game allows for exploratory learning by immersing the players inside a virtual heritage environment where they learn different aspects of history through their interactions with a crowd of virtual Roman avatars. The implementation of the Roma Nova game includes: (a) a crowd of Roman characters in the Forum and (b) a highly detailed set of buildings that belong to the Rome Reborn model. Intelligent agents are wandering around in the gaming environment between predefined points of interest, whereas the player is able to move freely and his movement is controlled via BCI device. To interact with the intelligent agents, the BCI-controlled player needs to approach them [11].

Participants and Experimental Procedure

Thirty-one participants used each of the prototypes. In case of both BCI systems the participants first trained using control panel (application provided by Emotiv SDK) and after that performed a task in virtual environment. After finishing the task, the participants were asked to evaluate their experience by filling in a questionnaire; a short unstructured interview also took place. These suggestions are a very helpful contribution towards the improvement of the system, giving feedback that an ordinary questionnaire cannot capture. The comparison of the results is presented in Table 6.1.

Results

Table 6.1 presents the comparison of user evaluation results of Roma Nova virtual environment and virtual robot with Emotiv headset. No significant differences for the ability to control, responsiveness, interaction and naturality of experience were found. The lack of significant differences can be explained by the similar difficulty

Table 6.1 Comparison of average rating values for Roma Nova virtual environment and virtual robot for Emotiv headset

Variable	Robot	Roma Nova	T-test(df)	Sig.
Ability to control	3.452	3.129	t(30) = 1.976	0.057
Responsiveness	3.226	3.581	t(30) = −1.688	0.102
Interaction	3.323	3.032	t(30) = 1.393	0.174
Naturality	3.484	3.290	t(30) = 0.862	0.395
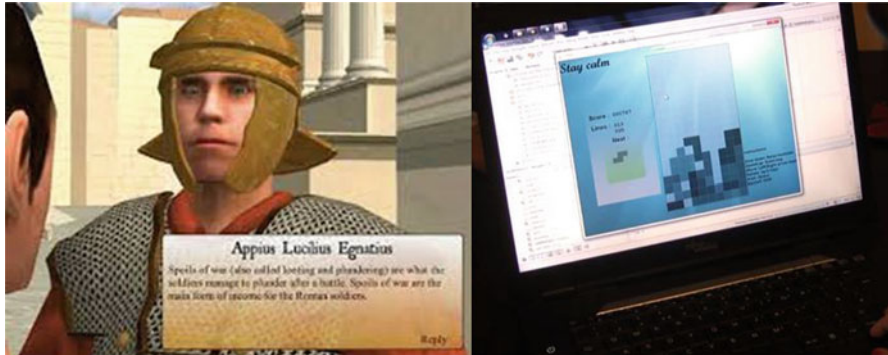
Fig. 6.3 BCI games used with NeuroSky device: *left* – Roma Nova, *right* – Tetris (Images used with permission from the Serious Games Institute, Coventry University)

of the BCI task. Both games required two-dimensional control, and while the quality of the virtual environment can influence the user experience it is unlikely to make a significant change.

Comparison of Serious Games Using NeuroSky

In this section we present the evaluation of two BCI systems which were developed with NeuroSky device; Roma Nova and Tetris (see Fig. 6.3).

Roma Nova

Roma Nova was used again to evaluate user experience while using the NeuroSky device. The participants were instructed to move to the particular point within the virtual environment. However, the main difference with the previous interaction paradigm is that only one sensor was used to fully control the avatar. In this case, the participants attempt to control the avatar by changing cognitive states such as meditation and attention, which is translated to two integer values (in the range 0–100). To turn right the participant has to concentrate as hard as possible, while in

order to move left they have to defocus their attention. Moving straight ahead was possible only by maintaining a balance between the two states. Meditation was used to control the velocity of the avatar, with high level of meditation resulting in high velocity.

Tetris

The second application use to evaluate NeuroSky was the well known Tetris game [13]. The Tetris game's purpose was to teach the players how to self-regulate their state of mind in a stressful demanding situation to their own benefit as the more meditative they manage to become the slower each shape will fall in the context of the levels speed. The difference this feature will make becomes more prominent as the levels increase. This serious game is a multi-threaded application where the speed of the current falling brick is determined by the number of milliseconds required for the shape to traverse one line down (move on the Y axis from y0 to y1). The bigger this step time value, the slower the brick will fall. Participants were asked to play the game three times with the end goal of scoring at least five lines each time. The speed of the falling shapes increased with each level and that a level was marked by the collapse of a line. The speed of the falling blocks was also dependent on the meditation level provided by the BCI device. The participants were given an unlimited training time in which they could get accustomed with the setting and rules of the game.

Participants and Experimental Procedure

31 participants used each of the prototypes. After completing each task, the users were asked to evaluate their experience using NASA TLX questionnaire [7], followed by a short unstructured interview also took place. The comparison of the results is presented in Table 6.2.

Results

Table 6.2 presents the comparison of user evaluation results of Roma Nova virtual environment and the Tetris game with NeuroSky headset. The users found controlling the avatar in Roma Nova virtual environment to be more mentally, physically and temporally demanding. They also reported that the Tetris game was less frustrating, required less effort, was easier to learn and the users scored the performance of the Tetris game higher than the performance of Roma Nova environment. There was no significant difference in terms of satisfaction gained by interaction with the two systems.

Table 6.2 Comparison of average rating values for Roma Nova virtual environment and Tetris game for NeuroSky headset

Variable	Roma Nova	Tetris	T-test (df)	Sig.
Mental demand	3.968	3.000	t(59) = 4,328	<0.001
Physical demand	4.032	1.933	t(59) = 9.198	<0.001
Temporal demand	2.516	2.667	t(59) = −0.531	=0.531
Performance[a]	2.452	3.933	t(55.144) = −5.602	<0.001
Effort	3.806	2.667	t(59) = 4,043	<0.001
Frustration	3.097	2.267	t(59) = 3.051	=0.003
Learnability	2.516	3.967	t(59) = −6.366	<0.001
Satisfaction	4.452	4.100	t(59) = 1.913	=0.61

[a]Due to the violation of the equality of variance assumption the result for equality of variance not assumed t-test is reported

Table 6.3 Comparison of average rating values between NuroSky and Emotiv headsets for games and virtual environments

Mann – Whitney U	z	P	NeuroSky	Emotiv	Sig.
Learnability	185.0	4.376	2.5161	3.6774	<0.001
Satisfaction	207.5	4.046	4.4516	3.4516	<0.001
Performance	211.5	3.957	2.4516	3.5806	<0.001
Effort	259.5	3.271	3.8065	3.5806	<0.01

Comparison of Emotiv and NeuroSky

The comparison between Nurosky and Emotiv devices was performed by [12]. In this case both devices were used to navigate an avatar in virtual environment of Roma Nova. The results are presented in Table 6.3 and they indicated that it is easier to control an avatar, achieved higher learnability and rated the performance higher when using Emotiv. Moreover, using Emotiv headset also required less effort than using NeuroSky. However, satisfaction was higher in Neurosky.

Conclusions and Limitations

The results presented in this chapter show that BCI technology is a viable option for use in serious games and virtual environments. The qualitative feedback provided by the users shows that they enjoyed the interactive experience, they were in favour of using EEG technology for interacting with games, even though it is not as accurate as joystick, computer mouse or keyboard. At this stage of development, its usability is however still limited. The results show no difference in user experience when an avatar or robot movement was controlled in 3D environment when a 14-channel device was used. The applications, which allow for control of an avatar in 3D

environment require two dimensional control (dimension one: forward – backward, dimension two: left – right). In case of one-channel device, the user's satisfaction was higher when the game used only one dimensional input (meditation level). Moreover, using one channel BCI device for the control of avatar in 3D environment results in lower user experience as compared to using BCI device with 14 channels.

The comparison of Emotiv and NeuroSky devices while using reach 3D environment shows that the 14-channel Emotiv device is better suited for this purpose. These results clearly indicate that the user experience was determined by the combination of the type of device used and the requirements of the game played. It is therefore important to match the requirements of the BCI application with the BCI device.

The investigations presented in this chapter were based only on self-reports provided by the users and the sample group included mostly students of computer science who may be biased when assessing new technology. Moreover, performing parametric statistical analysis on rating measurements, such as Likert scales gives rise to methodological problems. For more information on limitations of using measurements based on ratings in human-computer interaction see [31].

The potential of replacing the keyboard, computer mouse or joystick with a BCI device as the main channel of interaction for computer games seems far away. Substantial development in signal acquisition and signal precessing will have to be made to make this goal possible. In the mean time, many interesting and highly beneficial BCI applications can be proposed. In section "BCI Games for Medical Applications", we described a BCI game used for combating the symptoms of ADHD. BCI games for symptoms of other diseases can be imagined. One of them can be migraine and possibly epilepsy. Neuro-feedback based on slow cortical potentials regulation has been successfully user to stabilize brain activity and reduce the number of epileptic seizures [9] as well as to help control migraines [22]. It has, however, not yet been adapted for popular games to be used with commercially available BCI devices.

References

1. Baillet S, Mosher JC, Leahy RM (2001) Electromagnetic brain mapping. IEEE Signal Process Mag 18(6):14–30
2. Berger H (1969) On the electroencephalogram of man. Sixth report. Electroencephalogr Clin Neurophysiol Suppl–28:173
3. Birbaumer N (2006) Breaking the silence: brain–computer interfaces (BCI) for communication and motor control. Psychophysiology 43(6):517–532
4. Emotiv EEG Neuroheadset. www.emotiv.com/eeg/
5. Finke A, Lenhardt A, Ritter H (2009) The mindgame: a p300-based brain–computer interface game. Neural Netw 22(9):1329–1333
6. Guidi G, Frischer B, De Simone M, Cioci A, Spinetti A, Carosso L, Micoli LL, Russo M, Grasso T (2005) Virtualizing ancient Rome: 3D acquisition and modeling of a large plaster-of-Paris model of imperial Rome, Proc., Videometrics VIII, 56650D (17 Jan 2005). In: Electronic imaging 2005, San Jose. International Society for Optics and Photonics (SPIE 5665), pp 119–133. doi: 10.1117/12.587355

7. Hart SG, Staveland LE (1988) Development of nasa-tlx (task load index): results of empirical and theoretical research. Adv Psychol 52:139–183
8. Jackson MM, Mappus R, Barba E, Hussein S, Venkatesh G, Shastry C, Israeli A (2009) Continuous control paradigms for direct brain interfaces. In: Human-computer interaction. Novel interaction methods and techniques. Springer, Berlin/Heidelberg, pp 588–595
9. Kotchoubey B, Schneider D, Schleichert H, Strehl U, Uhlmann C, Blankenhorn V, Fröscher W, Birbaumer N (1996) Self-regulation of slow cortical potentials in epilepsy: a retrial with analysis of influencing factors. Epilepsy Res 25(3):269–276
10. Krepki R, Blankertz B, Curio G, Müller K-R (2007) The Berlin brain-computer interface (BBCI)–towards a new communication channel for online control in gaming applications. Multimed Tools Appl 33(1):73–90
11. Liarokapis F, Vourvopoulos A, Ene A, Petridis P (2013) Assessing brain-computer interfaces for controlling serious games. In: 2013 5th international conference on games and virtual worlds for serious applications (VS-GAMES). IEEE, Piscataway, pp 1–4
12. Liarokapis F, Debattista K, Vourvopoulos A, Petridis P, Ene A (2014) Comparing interaction techniques for serious games through brain–computer interfaces: a user perception evaluation study. Entertain Comput 5(4):391–399
13. Liarokapis F, Vourvopoulos A, Ene A (2015) Examining user experiences through a multi-modal BCI puzzle game. In: 2015 19th international conference on information visualisation (iV). IEEE, Piscataway, pp 488–493
14. Martin JH (1991) The collective electrical behavior of cortical neurons: the electroencephalo-gram and the mechanisms of epilepsy. Princ Neural Sci 3:777–791
15. Middendorf M, McMillan G, Calhoun G, Jones KS et al (2000) Brain-computer interfaces based on the steady-state visual-evoked response. IEEE Trans Rehabil Eng 8(2):211–214
16. Mindala. emotiv.com/store/apps/applications/132/12625
17. NeuroSky MindWave. http://store.neurosky.com/products/mindwave-1
18. Pope AT, Palsson OS (2001) Helping video games rewire "our minds". http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20040086464.pdf
19. Reuderink B, Nijholt A, Poel M (2009) Affective pacman: a frustrating game for brain-computer interface experiments. In: Intelligent technologies for interactive entertainment. Springer, Berlin/New York, pp 221–227
20. Roberts LE, Birbaumer N, Rockstroh B, Lutzenberger W, Elbert T (1989) Self-report during feedback regulation of slow cortical potentials. Psychophysiology 26(4):392–403
21. Shim B-S, Lee S-W, Shin J-H (2007) Implementation of a 3-dimensional game for developing balanced brainwave. In: 5th ACIS international conference on software engineering research, management & applications, 2007, SERA 2007. IEEE, Los Alamitos, pp 751–758
22. Siniatchkin M, Hierundar A, Kropp P, Kuhnert R, Gerber W-D, Stephani U (2000) Self-regulation of slow cortical potentials in children with migraine: an exploratory study. Appl Psychophysiol Biofeedback 25(1):13–32
23. StoneHange. http://emotiv.com/store/product_95.html
24. Strehl U, Leins U, Goth G, Klinger C, Hinterberger T, Birbaumer N (2006) Self-regulation of slow cortical potentials: a new treatment for children with attention-deficit/hyperactivity disorder. Pediatrics 118(5):e1530–e1540
25. Tan DS, Nijholt A (2010) Brain-computer interfaces: applying our minds to human-computer interaction. Springer, London
26. Tangermann M, Krauledat M, Grzeska K, Sagebaum M, Blankertz B, Vidaurre C, Müller K-R (2008) Playing pinball with non-invasive BCI. In: NIPS, Vancouver, pp 1641–1648
27. van de Laar B, Gurkok H, Plass-Oude Bos D, Poel M, Nijholt A (2013) Experiencing BCI control in a popular computer game. IEEE Trans Comput Intell AI Games 5(2):176–184
28. Vidal JJ (1977) Real-time detection of brain events in EEG. Proc IEEE 65(5):633–641
29. Vourvopoulos A, Liarokapis F (2014) Evaluation of commercial brain–computer interfaces in real and virtual world environment: a pilot study. Comput Electr Eng 40(2):714–729
30. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain–computer interfaces for communication and control. Clin Neurophysiol 113(6):767–791
31. Yannakakis GN, Martínez HP (2015) Ratings are overrated! Front ICT 2:13

Chapter 7
Psychophysiology in Games

Georgios N. Yannakakis, Hector P. Martinez, and Maurizio Garbarino

Abstract Psychophysiology is the study of the relationship between psychology and its physiological manifestations. That relationship is of particular importance for both game design and ultimately gameplaying. Players' psychophysiology offers a gateway towards a better understanding of playing behavior and experience. That knowledge can, in turn, be beneficial for the player as it allows designers to make better games for them; either explicitly by altering the game during play or implicitly during the game design process. This chapter argues for the importance of physiology for the investigation of player affect in games, reviews the current state of the art in sensor technology and outlines the key phases for the application of psychophysiology in games.

Introduction

Computer game players are presented with a wide and rich palette of affective stimuli during game play. Those stimuli vary from simple auditory and visual events (such as sound effects and textures) to complex narrative structures, virtual cinematographic views of the game world and emotively expressive game agents. Player emotional responses may, in turn, cause changes in the player's physiology, reflect on the player's facial expression, posture and speech, and alter the player's attention and focus level. Computer games, opposed to traditional music and video content, are highly interactive media that continuously react to the users' input. This interactivity can naturally accommodate mechanisms for real-time adaptation of game content aimed at adjusting player experience and realizing affective interaction [107].

The study of the relationship between psychology and its physiological manifestations defines the area of *psychophysiology* [15]. Physiology has been extensively

---

G.N. Yannakakis (✉) • H.P. Martinez
Institute of Digital Games, University of Malta, Msida, Malta
e-mail: georgios.yannakakis@um.edu.mt; hector.p.martinez@um.edu.mt

M. Garbarino
Empatica, Milan, Italy
e-mail: mg@empatica.com

investigated in relation to affect ([3, 19] among many others) so the relationship between physiology and affect is by now undeniable; the exact mapping, however, is still far from known. What is widely evidenced is that the sympathetic and the parasympathetic components of the autonomic nervous system are involuntary affected by affective stimuli. In general, arousal-intense events cause dynamic changes in both nervous systems: an increase and a decrease of activity, respectively, at the sympathetic and the parasympathetic nervous system. Alternatively, activity at the parasympathetic nervous system is high during relaxing or resting activities. In turn, such nervous system activities cause alterations in one's electrodermal activity, heart rate variability, blood pressure, and pupil dilation [15, 88].

This relation between physiology and affect has been exploited in game research to detect player affect [106]. While some studies have investigated physiological reactions in isolation, researchers often look at the reactions to aspects of the game context [61, 65, 79]. The context of the game during the interaction is a necessary input for appropriately detecting the psychophysiological responses of players. The game context—naturally fused with other input modalities from the player—has been used in several studies to predict different affective states and other dissimilar mental states relevant to playing experience ([71, 83, 86] among others). The fusion of physiology and gameplay or player behavioral metrics has been explored in a small number of studies, typically by analyzing the physiological responses to game events [20, 38, 79] but also using physiological and gameplay statistical features [60, 66]. Other modalities that have been explored extensively but are covered in other parts of this book include facial expressions [4, 14, 36, 49, 111], muscle activation (typically face) [20, 26], body movement and posture [5, 11, 28, 49, 96], speech [7, 43, 45, 47, 97], brain interfaces [1, 81] and eye movement [5].

At the moment of writing there are a few examples of commercial games that utilize physiological input from players. Most notably *Nevermind* (Flying Mollusk 2015) is a biofeedback-enhanced adventure horror game that adapts to the player's stress levels by increasing the level of challenge it provides: the higher the stress the more the challenge. A number of sensors are available for affective interaction with *Nevermind* which include skin conductance and heart activity. *The Journey of Wild Divine* (Wild Divine 2001) is another biofeedback-based game designed to teach relaxation exercises via the player's blood volume pulse and skin conductance. It is also worth noting that AAA game developers such as Valve have already experimented with the player's physiological input for the personalization of games such as *Left 4 Dead* (Valve 2008) [2].

This chapter builds upon the important association between player experience and physiology in games, it provides a quick guide on the sensor technology available, and it outlines the key phases for building effective physiology-based affective interaction in games: annotation, modeling, and adaptation. The chapter explicitly excludes electroencephalography (EEG) from the physiological signals covered; EEG defines the core topic of another chapter of this book.

Why Physiology in Games?

Arguably several modalities of player input are still nowadays *implausible* within commercial-standard game development. Pupillometry and gaze tracking are very sensitive to distance from screen and variations in light and screen luminance, which makes them rather impractical for use in a game application. Camera-based modalities (facial expressions, body posture and eye movement) require a well-lit environment often not present in home settings (e.g. when playing video-games) and they can be seen by some users as privacy hazards (as the user is continuously recorded). Even though highly unobtrusive the majority of the vision-based affect-detection systems currently available cannot operate well in real-time [111]. Speech is a highly accessible, real-time efficient and unobtrusive modality with great potential for gaming applications (see corresponding chapter on speech); however, it is only applicable to either games where speech forms a control modality (as e.g. in conversational games for children [48, 110]) or collaborative games that naturally rely on speech for communication across players (e.g. in collaborative first person shooters). Aside the potential they might have, the appropriateness of facial expression, head pose and speech for emotion recognition in games is questionable since experienced players tend to stay still and speechless while playing games [6]. Further details about affect detection in games via images, videos and speech are given in other chapters of this book.

Recent years have seen a significant volume of studies that explore the interplay between physiology and gameplay by investigating the impact of different gameplay stimuli to dissimilar physiological signals ([29, 58, 59, 65, 68, 78, 90, 95] among others). Such signals are usually obtained through electrocardiography (ECG) [109], photoplethysmography [95, 109], galvanic skin response (GSR) [39–41, 59], respiration [95], EEG [70] and electromyography (EMG).

Existing hardware for EEG, respiration and EMG require the placement of body parts such as the head, the chest or parts of the face to the sensors making those physiological signals rather impractical and highly intrusive for most games. On the contrary, recent sensor technology advancements for the measurement of electrodermal activity (skin conductivity), photoplethysmography (blood volume pulse), heart rate variability and skin temperature have made those physiological signals even more attractive for the study of affect in games. Real-time recordings of these can nowadays be obtained via comfortable wristbands and stored in a personal computer or a mobile device via a wireless connection.

It is evident that we can measure physiological responses to external stimuli via several modalities of player input. Due to space constraints, however, in this chapter we focus primarily on the two most popular, real-time efficient and appropriate signals for affective games: electrodermal activity and heart activity. Before delving into the details of the sensor technology available and the methods for modeling player's affect via physiology we herein outline the key properties of these two core physiological signals and their importance for psychophysiological studies (in games and beyond).

Heart Activity

Heart rate variability (HRV) refers to the physiological phenomenon that causes variation in the time window between consequent heartbeats. HRV and heart rate are derived through the detection of heart beats. The two core methods used to detect heart beats include the electrocardiogram (ECG) and the pulse wave signal derived from a photoplethysmograph (PPG)—also known as blood volume pulse sensor. While ECG is generally considered superior compared to blood volume pulse (as it provides a much clearer signal) it is not practical for affective gaming applications since it requires the use of electrodes placed on a player's chest.

There are numerous studies suggesting that heart rate and HRV are associated with emotional arousal. In particular, the high-frequency (HF) band of HRV activity has been found to decrease with elevated anxiety [44]. On that basis, HRV has been shown to be reduced under reported stress and worry states [13]. Moreover, it has been suggested that the HF band of HRV is mainly driven by respiration and appears to derive mainly from vagal activity [35]. Specifically, the energy of the HF range, representing quicker changes in heart rate, is primarily due to parasympathetic activity of the heart which is decreased during mental or stress load [35]. The multimodal association of heart rate and HRV to emotion and the real-time efficiency of available HRV sensors have made it a very popular measure of emotive activity in games (see [41, 101] among many).

Electrodermal Activity

Electrodermal activity (EDA) is the ability of the human body to cause continuous variation in the electrical characteristics of the skin [12]. EDA is a core bodily response when the sympathetic branch of the autonomic nervous system is activated due to a stimulus. What is unique about the human skin is that is the only organ that responds solely to alterations of the sympathetic nervous system; skin is not affected by activities on the parasympathetic nervous system. Essentially, an external or internal stimulus may activate the sympathetic nervous system, which in turn, activates the glands to release sweat. Sweat yields increased electrical activity which can be detected via electrical potentials between electrodes placed on the skin. These electrodes are usually placed on the fingertips, the toes or the wrist.

The direct relationship between EDA and sympathetic arousal is well researched and evidenced by now. As a result EDA is the most popular method for investigating human psychophysiological phenomena [12] and skin conductivity is currently amongst the most common modalities for measuring emotive responses that are associated to arousal such as stress, frustration and anxiety (see [40, 41, 79, 101] among many). Beyond affect, EDA has also been associated with manifestations of cognitive processes [24].

Sensor Technology

Physiological sensor technology has seen significant advancements over the last decade. The 8-channel *ProComp Infiniti*[1] (see Fig. 7.1c) was among the first hardware devices used broadly for research in psychophysiology in e.g. physical interactive games [101, 108] (blood volume pulse and skin conductance), and racing games [95] (respiration, blood volume pulse, skin conductance, and skin temperature). While providing signals of clinical-standard resolution, the *ProComp Infiniti* device proved to be cumbersome for use in games due to its sensitivity to movements and impractical for broad use due to its cost. In addition, all aforementioned studies report the significant technical challenges faced with the blood volume pulse sensor and its placement. Due to the lack of a grip for appropriate attachment to a finger or ear lobe (see Fig. 7.1c), the BVP sensor yielded noise-enhanced signals that were challenging to process, to extract features from and/or to derive the heart rate and heart rate variability of the player. Some other popular devices for measuring skin conductance and/or heart activity include the Biopac GSR100C [10], the Affectiva's[2] Q Sensor (which is no longer available), the BodyMedia Sensewear [52], the BodyBugg armband and the Nymi band.[3] All above devices, however, have seen very limited use in gaming applications as they (a) do not allow access to real-time data (BodyMedia Sensewear, BodyBugg, Nymi), (b) are highly intrusive (the Boipac device requires the application of conductive gel), (c) they are very sensitive to movement (Q sensor, Biopac), or (d) they are very expensive for broad gaming applications (e.g. Q sensor, ProComp Infiniti).

In recent years physiological sensor technology has delivered a plethora of sensors that—compared to the aforementioned devices—are both more reliable for data collection and more appropriate for gaming applications. A notable example is the IOM biofeedback device which consists of three sensors: two electrodes for skin conductance and one blood volume pulse sensor placed on the subject's fingertips (see Fig. 7.1a). The use of small and accurate commercial apparatus like the IOM biofeedback device in the least intrusive way minimizes (psychological) experiment effects caused by the presence of recording devices and maximizes data reliability. For its real-time efficiency, low cost and good data quality—mainly due to the robust finger grips of the sensors—IOM has been used extensively in several studies for psychophysiology in games (e.g. see [40, 41, 109] among many). Furthermore, IOM is the key sensor for commercial biofeedback games such as *Nevermind* (Flying Mollusk 2015) and *The Journey of Wild Divine* (Wild Divine 2001). Another example of a successful wearable sensor is the Empatica's[4] *Embrace* wristband (see Fig. 7.1b). Embrace is built on the technical know how of the E4

---

[1] http://thoughttechnology.com/

[2] http://www.affectiva.com

[3] https://www.nymi.com/

[4] http://www.empatica.com/

Fig. 7.1 The key physiological signal sensors and devices discussed in this chapter. (**a**) The IOM device used during the data collection experiment reported in [109]. (**b**) Empatica's *Embrace* wristband. (**c**) The blood volume pulse sensor of the *ProComp Infiniti* device. (**d**) The *Cardiio* application for smartphones

wristband (used e.g. in [41]) and the Q sensor and measures skin conductivity, skin temperature and 3D movement (via an accelerometer and a gyroscope). It is real-time efficient, highly unobtrusive for both home gaming settings and mobile gaming *in the wild* and provides more reliable data compared to earlier wrist-based devices.

Nowadays there are quite a few smartphone/tablet software applications that are able to support camera-based pulse detection (contact-less physiological measurement) such as the *Strees Check* app for Android by Azumio; however, access to real-time HRV data is not available to the user in most of these apps (if not all). We particularly note the heart rate *Cardiio* app[5] (see Fig. 7.1d) which is build on early studies on face-based pulse detection [76]. Cardiio approximates heart rate through the face's light reflection which is affected by the amount of blood

---

[5]http://www.cardiio.com/

available on a face. A heart beat increases the amount of blood into one's face which results in lower levels of light reflection. Measurement accuracy for all these mobile applications is very close (e.g. up to a 3 beat per minute difference) to a clinical pulse oximeter; however, the data is reliable only when the mobile's or the tablet's camera is used in a well-lit environment.

For an extensive discussion on available physiological sensors and their corresponding strengths and weaknesses the interested reader may refer to [88].

Annotating Physiology with Psychological Labels

The question of how to best annotate affect has been a milestone challenge for affective computing. Appropriate methods and tools addressing that question can provide better estimations of the ground truth which, in turn, may lead to more efficient affect detection and more reliable models of affect. Affect annotation becomes even more challenging within games due to their fast-paced and rich affective interaction.

Manually annotating emotion in games is a challenge in its own right both with respect to the human annotators involved and the annotation protocol chosen. On one hand, the human annotators need to be skilled enough to be able to approximate the perceived affect well and, therefore, eliminate subjective biases introduced to the annotation data. On the other hand, there are many open questions left for the designer of the annotation study when it comes to the annotation tools and protocols used. Will the person experiencing the emotion (first person) or others (third-person) do the labeling? How well trained (or experienced) should the annotators be and how will the training be done? Will the labeling of emotion involve states (discrete representation) or does it involve the use of emotion intensity or affect dimensions (continuous representation)? When it comes to time, should it be done in real-time or offline, in discrete time periods or continuously? Should the annotators be asked to *rate* the affect in an absolute fashion or, instead, *rank* it in a relative fashion? Answers to the above questions yield different data annotation protocols and, inevitably, data quality, validity and reliability.

Representing both time and emotion as a continuous function has been one of the dominant annotation practices within affective computing over the last 15 years. Continuous labeling *with respect to emotion* appears to be advantageous compared to discrete states labeling for several reasons. The states that occur in naturalistic data hardly fit word labels or linguistic expressions with fuzzy boundaries. Further, when states are used it is not trivial to capture variations in emotion intensity and, as a result, earlier studies have shown that inter-rater agreement tends to be rather low [21]. The dominant approach in continuous annotation is the use of Russell's two-dimensional (arousal-valence) circumplex model of affect [84]. Valence refers to how pleasurable (positive) or unpleasurable (negative) the emotion is whereas arousal refers to how intense (active) or lethargic (inactive) that emotion is.

Continuous labeling *with respect to time* has been popularized due to the existence of tools such as FeelTrace (and its variant GTrace [22]) which is a freely available software that allows real-time emotional annotation of video content [23], the continuous measurement system [67] which has also been used for annotating videos, and EmuJoy [69] which is designed for the annotation of music content. The real-time continuous annotation process, however, appears to require a higher amount of cognitive load compared to e.g. offline and discrete annotation protocols. Such cognitive load often results in low inter-rater agreement and unreliable data annotation [27, 57].

The most direct way to annotate an emotion in games is to ask the players them-selves about their playing experience and build a model based on these annotations. Subjective emotion annotation can be based on either players free-response during play (think aloud protocols) or on forced data retrieved through questionnaires. Alternatively, experts or external observers may annotate the playing experience in a similar fashion. Third-person emotion annotation entails the identification of particular affective states by user experience and game design experts. The annotation is usually based on the triangulation of multiple modalities of player and game input such as the players head pose, in-game behaviour and game context [87].

Annotations (either self-reports or third-person) can be classified as rating (scalar), class and preference. In rating, annotators are asked to answer questionnaire items given in a rating/scaling form (as in [59])—such as the affective aspects of the Game Experience Questionnaire [75]—which labels affective states with a scalar value (or a vector of values). In a class-based format subjects are asked to pick an affective state from a particular representation which could vary from a simple boolean question (was that game level frustrating or not? is this a sad facial expression?) to an affective state selection from e.g. the Geneva Emotion Wheel [8]. Finally, subjects are able to provide answers in a rank-based (preference) format, in which they are asked to compare an affective experience in two or more variants/sessions of the game ([99] among others) (was that level more engaging that this level? Which facial expression looks happier?). A plethora of recent studies in the area of affective annotation in games (and beyond) [40, 63, 95, 98, 103–105] have shown the supremacy of rank-based emotion annotation over rating and class-based annotation.

Models of Psychophysiology in Games

In this section we outline the key phases of modeling physiological responses which are labeled with affect annotations—i.e. deriving the mapping between player affect and its physiological manifestations—and the challenges games pose to each one of these phases. The phases we describe follow the core affect detection steps [17] and include signal processing, feature extraction and selection, and modeling.

Physiological Signal Processing

Physiological signals are unidimensional time series, the quality and reliability of which is dependent on the sensor technology available and the experiment protocol followed. In that regard the signals are subject to standard preprocessing and noise removal methods. Popular techniques include wavelet transform thresholding and least mean square adaptive filters [37].

Games pose additional challenges when it comes to data collection via physiological signals. First, particular sensors such as EEG or electrocardiogram can be highly intrusive which, in turn, affects the quality of play and data gathered. Second, the interaction in games is fast-paced and rich causing rapid body movements and quick alterations in emotive states. Finally, there are so many factors contributing to player experience (and affecting it) that not even the most carefully designed controlled experiment can eliminate the potential effects manifested through a player's physiology. For an extensive overview of techniques for data preprocessing on physiological signals one may refer to [16].

Feature Extraction

Once data is denoised any feature extraction mechanism is applicable to the signals. Examples of feature extraction methods include standard ad-hoc (manual) feature extraction such as average and standard deviation values of the signal, principal component analysis and Fisher's linear discriminant analysis. Focusing on the particularity of skin conductance as a signal for feature extraction it is worth noting that the trough-to-peak analysis of galvanic skin response can be subject to superpositioning of phasic and tonic activity. This necessitates the subtraction of baseline measures or other forms of signal correction [12]. It has been suggested that even with such corrections one may still confound phasic and tonic skin conductance [9] which is undesirable in games as they predominantly activate skin conductance via particular in-game events. To address this issue, features of a player's skin conductance can be extracted using continuous decomposition analysis [9]. The method allows for the decomposition of phasic and tonic electrodermal activity and has been applied for stress detection in games [40].

Physiological feature extraction is naturally enriched through the game context. To this end important game events can be used to determine the response time window that features can be extracted from. A number of studies have been adopting this event-based feature extraction approach for variant psychological signals [40, 50, 61, 79, 80].

Because of the rich affective interaction and the availability of multitude types and amounts of emotion elicitors, physiological signals derived from games are rather complex to extract relevant features from. While standard methods used in affective computing might suffice evidence in the literature has shown that methods

such as sequence mining [61] and deep learning [62] yield richer representations of affect manifestations in games. In the study of Martinez and Yannakakis [61] frequent subsequent physiological manifestations are fused with in game events to provide relevant features for affect modeling. In the study of Martinez et al. [62], deep learning can derive more complex temporal signal features that yield higher affect model accuracies compared to standard (ad-hoc) designed features.

Feature Selection

Once features are extracted the subset of the most relevant features for a particular affective state or emotion dimension (e.g. arousal) need to be derived from the set of features available. It is desired that the affective model constructed is dependent on a minimal number of features that yield the highest prediction accuracy. The primary reasons for minimizing the feature subset are improvements of model expressiveness (interpretability) and reduction of computational effort in training and real-time performance. Therefore, feature selection is utilized to find the feature subset that yields that most accurate affective model and save computational effort of exhaustive search on all possible feature combinations. The quality of the predictive model constructed (see next subsection) depends critically on the set of input data features chosen. The resulting set of physiological features define the input to the affect model. Studies within affective games have so far primarily used sequential forward selection, sequential backward selection and genetic search-based feature selection [60, 99].

Modeling Psychophysiology

A model of a player's psychophysiology predicts some aspect of the experience of a player in general, a type of player or a particular player would have in some game situation. If data recorded includes a scalar representation of affect, or classes and annotated labels of user states, any of a large number of machine learning (regression and classification) algorithms can be used to build affective models. Available methods include neural networks, Bayesian networks, decision trees, support vector machines and standard linear regression. On the other hand, if the ground truth of player experience is given in a pairwise preference (rank) format (e.g. game version X is more frustrating than game version Y) standard supervised learning techniques are inapplicable, as the problem becomes one of *preference learning* [33, 99]. Available preference learning approaches include linear discriminant analysis, decision trees, artificial neural networks (shallow and

deep architectures) and support vector machines. A number of such methods are currently included in the open-access Preference Learning Toolbox[6] [32].

Adapting the Game to Affect Models

For affective interaction to be realized the game logic needs to adapt to the current state of the game-player interaction. Whether agent behavior or parameterized game content, a mapping is required linking a user's affective state to the game context. That mapping is essentially the outcome of the emotion modeling phase described above. Any search algorithm (varying from local and global search to metaheuristic and exhaustive search) is applicable for searching in the parameterised search space and finding particular game states (context) that are appropriate for a particular affective state of a specific player. For example, one can envisage the optimization of agent behavior attributes for maximizing engagement, frustration or empathy towards a player [51]. As another example, the study of Shaker et al. [86] presents the application of exhaustive search for generating *Super Mario Bros* (Nintendo 1985) levels that are maximally frustrating, engaging or challenging for any player.

There are a number of elements (i.e. game content) from the game world that an adaptive process can alter in order to drive the player to particular affective patterns. Game content may include every aspect of the game design such as game rules [91], reward systems, lighting [25, 31], camera profiles [109], maps [93], levels [86], tracks [92, 94], story plot points [34, 82], sound [55, 56] and music [30]. Even behavioral patterns of NPCs such as their navigation meshes, their parameterized action space and their animations can be viewed as content. The adaptive process in this case is referred to as *procedural content generation* (PCG) which is the generation of game content via the use of algorithmic means. According to the taxonomy presented in [94] game content can be *necessary* (e.g. game rules) or *optional* (e.g. trees in a level or flying birds on the background). Further, PCG can be either *offline* or *online*, *random* or based on a *parameterised space*, *stochastic* or *deterministic* and finally it can be either *constructive* (i.e. content is generated once) or *generate-and-test* (i.e. content is generated and tested). The experience-driven PCG framework [107] views game content as an indirect building block of player affect and proposes adaptive mechanisms for synthesizing personalised game experiences.

A critical question once an adaptation mechanism is designed is how often particular attributes should be adjusted. The frequency can vary from simple predetermined or dynamic time windows [102] but adaptation can also be activated every time a new level [86] or a new game [100] starts, or even after a set of critical player actions—such as in *Façade* [64]. The time window of adaptation is heavily dependent on the game under examination and the desires of the game designer.

---

[6]http://sourceforge.net/projects/pl-toolbox/

Regardless of the time window adopted, adaptation needs to be interwoven well with design if is to be successful.

Psychophysiology Beyond Games

In this section we argue for the broad impact of psychophysiological research and we identify and briefly survey two primary application domains: education (via intelligent tutoring systems) and health. While games have been used extensively in both of these domains simpler modes of human computer interaction (such as mere simulations of virtual agents or tutors) are more common.

Intelligent Tutoring Systems

Confusion, anxiety and frustration are cognitive and affective states with a direct impact on students' learning process and outcome [74, 85]. Consequently, affect detection has become increasingly important in the intelligent tutoring systems (ITS) community [83]. The core idea is to enhance the learning capacity of a student and the learning experience (via e.g. minimizing frustration) through a virtual (intelligent) tutor that is capable of detecting the affective state of the student and reacting to it. Research in ITS has mostly focused on the detection phase [18], evaluating dissimilar methods to model student confusion [36, 42], frustration [20, 65] and attention [77]. An example of game-based virtual tutors that react to automatically detected affect can be found in [83]. Even when tutoring systems are not realized through games, one can argue that a learning activity via interaction with a virtual tutor and a learning activity through a game-based scenario yield similar psychophysiological patterns. As a consequence, the methodology covered throughout this chapter is directly relevant for the study of intelligent tutoring systems.

Health Technologies

Nowadays, a significant part of the world's population is afflicted by depression and anxiety-related disorders, which are directly connected to emotion and moods. Affect detection can be the key for the diagnosis and computer-based treatment of such mental health issues. Post-traumatic stress disorder (PTSD) has attracted a lot of attention within the affective computing literature. Holmgaard et al. [40, 41] have conducted representative research in this area. They designed and developed a game-based tool for treating PTSD based on stress inoculation and exposure therapy techniques. Physiological signals such as galvanic skin response and HRV were

recorded from patients. Those signals would be processed to derive stress profiles for the patient based on his skin conductance manifestations of stress on particular in-game auditory and visual events. Those stress profiles can be used both as a diagnostic and as an assistive tool for PTSD (see more details in the games for health chapter of this book).

Another application of affect detection to health technologies is related to syndromes such as autism that involve difficulties processing or expressing emotions. There has been a large body of studies in affective computing research towards developing tools to help parents, teachers and carers of children with autism [46, 54, 72]. These tools detect the affective state of children and communicate it to themselves or others, enhancing communication.

An additional application of psychophysiology to health technologies has been explored in relation to tele-medicine. In this particular domain, emotion is not at the core of the treated illness but it is regarded as an important element of the communication between the patient and the doctor. Detecting the affective state of the patient can help the doctor to better diagnose or simply better interact with the patient. This enhanced communication can improve the patient's satisfaction and lead to a faster recovery. Lisetti et al. [53] developed such a system, in which the affective state of the patient was predicted from her physiological signals and directly communicated to the doctor.

Limitations of Physiology

As already mentioned, most existing hardware for physiological recording require the contact of body parts (e.g. head, chest or fingertips) to sensors making physiological signals such as EEG and respiration rather impractical and highly intrusive. Furthermore some sensors are still very costly for a broad use in gaming. As seen in section "Sensor Technology", however, recent advances in sensor technology have resulted in low-cost unobtrusive biofeedback devices appropriate for gaming applications (such as the IOM and the *Embrace* wristband). In addition, contact-less heart activity detection applications such as *Cardiio* offer a promising future for physiology-based gaming.

Another point of concern for the use of physiology-based game interaction is the effect of signal habituation. Habituation is the learning process of the autonomic nervous system when exposed to a particular stimulus several times. According to Solokov [89] the nervous system creates a representation (model) of the stimulus which is updated each time the stimulus is presented. The closer the expected representation (model) comes to the actual stimulus the lower the affect to bodily responses, which in turn yield physiological habituation. Habituation is of particular relation to game-related research and connected to learnability in games. The design of a successful game-based affective interaction approach should be able to provide dissimilar stimuli or control for habituation.

Physiological responses are affected by numerous factors including mood, physical movement, physical state, age, blood sugar levels, caffeine consumption, and drug use. To eliminate as many subjectivity biases as possible one needs to record the physiological state of a subject during a short resting period prior to any gameplay session. Baseline recordings from that period shall be used to both offset the signals prior to affect modeling and calibrate any resulting affect models during the interaction [73].

Conclusions

This chapter explored the potential of psychophysiology in gaming applications and argued for the importance of physiology for achieving affective interaction and enhanced player experience. Putting an emphasis on heart and electrodermal activity we surveyed the current state of the art in sensor technology and outlined the key phases of physiology-based affect detection and modeling. We also discussed the evident potential of psychophysiology (through games or other applications) in domains such as intelligent tutoring systems and health. Existing studies in the literature, available sensor technology and a plethora of commercial-standard games that incorporate psychophysiological processes as game affordances suggest that physiology is an important means for realizing affective interaction in games with a great potential for further research and development.

References

1. AlZoubi O, Calvo R, Stevens R (2009) Classification of EEG for affect recognition: an adaptive approach. In: AI 2009: advances in artificial intelligence. Springer, pp 52–61
2. Ambinder M (2011) Biofeedback in gameplay: how Valve measures physiology to enhance gaming experience. In: Game developers conference, San Francisco
3. Andreassi JL (2000) Psychophysiology: human behavior and physiological response. Psychology Press
4. Arroyo I, Cooper DG, Burleson W, Woolf BP, Muldner K, Christopherson R (2009) Emotion sensors go to school. In: Proceedings of conference on artificial intelligence in education (AIED). IOS Press, pp 17–24
5. Asteriadis S, Tzouveli P, Karpouzis K, Kollias S (2009) Estimation of behavioral user state based on eye gaze and head pose–application in an e-learning environment. Multimed Tools Appl 41(3):469–493
6. Asteriadis S, Karpouzis K, Shaker N, Yannakakis GN (2012) Does your profile say it all? Using demographics to predict expressive head movement during gameplay. In: UMAP workshops, citeseer
7. Banse R, Scherer KR (1996) Acoustic profiles in vocal emotion expression. J Personal Soc Psychol 70(3):614

8. Bänziger T, Tran V, Scherer KR (2005) The Geneva emotion wheel: a tool for the verbal report of emotional reactions. Poster presented at ISRE

9. Benedek M, Kaernbach C (2010) A continuous measure of phasic electrodermal activity. J Neurosci Methods 190(1):80–91

10. Bersak D, McDarby G, Augenblick N, McDarby P, McDonnell D, McDonald B, Karkun R (2001) Intelligent biofeedback using an immersive competitive environment. Paper at the designing ubiquitous computing games workshop at UbiComp

11. Bianchi-Berthouze N, Lisetti CL (2002) Modeling multimodal expression of users affective subjective experience. User Model User-Adapt Interact 12(1):49–84

12. Boucsein W (2012) Electrodermal activity. Springer, New York

13. Brosschot JF, Van Dijk E, Thayer JF (2007) Daily worry is related to low heart rate variability during waking and the subsequent nocturnal sleep period. Int J Psychophysiol 63(1):39–47

14. Busso C, Deng Z, Yildirim S, Bulut M, Lee CM, Kazemzadeh A, Lee S, Neumann U, Narayanan S (2004) Analysis of emotion recognition using facial expressions, speech and multimodal information. In: Proceedings of international conference on multimodal interfaces (ICMI). ACM, pp 205–211

15. Cacioppo JT, Berntson GG, Larsen JT, Poehlmann KM, Ito TA et al (2000) The psychophysiology of emotion. In: Lewis M, Haviland-Jones JM (eds) Handbook of emotions, vol 2. Guilford Press, New York, pp 173–191

16. Cacioppo JT, Tassinary LG, Berntson G (2007) Handbook of psychophysiology. Cambridge University Press, Cambridge/New York

17. Calvo RA, D'Mello S (2010) Affect detection: an interdisciplinary review of models, methods, and their applications. IEEE Trans Affect Comput 1(1):18–37

18. Calvo RA, D'Mello SK (2011) New perspectives on affect and learning technologies, vol 3. Springer, New York

19. Calvo R, Brown I, Scheding S (2009) Effect of experimental factors on the recognition of affective mental states through physiological measures. In: AI 2009: advances in artificial intelligence. Springer, pp 62–70

20. Conati C, Maclaren H (2009) Modeling user affect from causes and effects. In: User modeling, adaptation, and personalization, Trento, pp 4–15

21. Cowie R, Cornelius RR (2003) Describing the emotional states that are expressed in speech. Speech Commun 40(1):5–32

22. Cowie R, Sawey M (2011) Gtrace-general trace program from queens, belfast

23. Cowie R, Douglas-Cowie E, Savvidou S, McMahon E, Sawey M, Schröder M (2000) 'FEELTRACE': an instrument for recording perceived emotion in real time. In: ISCA tutorial and research workshop (ITRW) on speech and emotion

24. Critchley HD, Mathias CJ, Dolan RJ (2002) Fear conditioning in humans: the influence of awareness and autonomic arousal on functional neuroanatomy. Neuron 33(4):653–663

25. De Melo C, Paiva A (2007) Expression of emotions in virtual humans using lights, shadows, composition and filters. In: Affective computing and intelligent interaction. Springer, Berlin/New York, pp 546–557

26. Dennerlein J, Becker T, Johnson P, Reynolds C, Picard RW (2003) Frustrating computer users increases exposure to physical factors. In: Proceedings of the international ergonomics association (IEA), Seoul

27. Devillers L, Vidrascu L (2006) Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs. In: Proceedings of conference of the international speech communication association (Interspeech), Pittsburgh, pp 801–804

28. D'Mello S, Graesser A (2009) Automatic detection of learner's affect from gross body language. Appl Artif Intell 23(2):123–150

29. Drachen A, Nacke L, Yannakakis GN, Pedersen AL (2010) Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In: Proceedings of the SIGGRAPH symposium on video games. ACM-SIGGRAPH Publishers, New York

30. Eladhari M, Nieuwdorp R, Fridenfalk M (2006) The soundtrack of your mind: mind music-adaptive audio for game characters. In: Proceedings of the 2006 ACM SIGCHI international conference on advances in computer entertainment technology. ACM, p 54

31. El-Nasr MS, Vasilakos A, Rao C, Zupko J (2009) Dynamic intelligent lighting for directing visual attention in interactive 3-D scenes. IEEE Trans Comput Intell AI Games 1(2):145–153
32. Farrugia VE, Martínez HP, Yannakakis GN (2015) The preference learning toolbox. arXiv preprint arXiv:1506.01709
33. Fürnkranz J, Hüllermeier E (2005) Preference learning. Künstliche Intelligenz 19(1):60–61
34. Giannatos S, Nelson M, Cheong Y-G, Yannakakis GN (2011) Suggesting new plot elements for an interactive story. In: Proceedings of the 4th workshop on intelligent narrative technologies, AIIDE, AAAI Press
35. Goldberger JJ, Challapalli S, Tung R, Parker MA, Kadish AH (2001) Relationship of heart rate variability to parasympathetic effect. Circulation 103(15):1977–1983
36. Grafsgaard J, Boyer K, Lester J (2011) Predicting facial indicators of confusion with hidden Markov models. In: Proceedings of international conference on affective computing and intelligent interaction (ACII). Springer, Memphis, pp 97–106
37. Haykin S, Widrow B (2003) Least-mean-square adaptive filters, vol 31. Wiley, Hoboken
38. Hazlett RL (2006) Measuring emotional valence during interactive experiences: boys at video game play. In: Proceedings of SIGCHI conference on human factors in computing systems (CHI). ACM, New York, pp 1023–1026
39. Holmgård C, Yannakakis GN, Karstoft K-I, Andersen HS (2013) Stress detection for PTSD via the startlemart game. In: 2013 humaine association conference on affective computing and intelligent interaction (ACII). IEEE, Piscataway, pp 523–528
40. Holmgård C, Yannakakis GN, Martínez HP, Karstoft K-I (2015) To rank or to classify? Annotating stress for reliable PTSD profiling. In: 2015 international conference on affective computing and intelligent interaction (ACII), Xi'an
41. Holmgård C, Yannakakis GN, Martínez HP, Karstoft K-I, Andersen HS (2015) Multimodal PTSD characterization via the startlemart game. J Multimodal User Interfaces 9(1):3–15
42. Hussain M, AlZoubi O, Calvo R, D'Mello S (2011) Affect detection from multichannel physiology during learning sessions with autotutor. In: Proceedings of international conference in artificial intelligence in education (AIED). Springer, Heidelberg, pp 131–138
43. Johnstone T, Scherer KR (2000) Vocal communication of emotion. In: Lewis M, Haviland-Jones JM (eds) Handbook of emotions, vol 2. Guilford Press, New York, pp 220–235
44. Jönsson P (2007) Respiratory sinus arrhythmia as a function of state anxiety in healthy individuals. Int J Psychophysiol 63(1):48–54
45. Juslin PN, Scherer KR (2005) Vocal expression of affect. Oxford University Press, Oxford
46. Kaliouby R, Picard R, Baron-Cohen S (2006) Affective computing and autism. Ann N Y Acad Sci 1093(1):228–248
47. Kannetis T, Potamianos A (2009) Towards adapting fantasy, curiosity and challenge in multimodal dialogue systems for preschoolers. In: Proceedings of international conference on multimodal interfaces (ICMI). ACM, New York, pp 39–46
48. Kannetis T, Potamianos A, Yannakakis GN (2009) Fantasy, curiosity and challenge as adaptation indicators in multimodal dialogue systems for preschoolers. In: Proceedings of the 2nd workshop on child, computer and interaction. ACM, New York, p 1
49. Kapoor A, Burleson W, Picard RW (2007) Automatic prediction of frustration. Int J Hum-Comput Stud 65(8):724–736
50. Kivikangas JM, Ekman I, Chanel G, Järvelä S, Salminen M, Cowley B, Henttonen P, Ravaja N (2010) Review on psychophysiological methods in game research. In: Procedings of Nordic digital games research association conference (Nordic DiGRA)
51. Leite I, Mascarenhas S, Pereira A, Martinho C, Prada R, Paiva A (2010) "Why can't we be friends?" An empathic game companion for long-term interaction. In: Intelligent virtual agents. Springer, Berlin, pp 315–321
52. Lisetti CL, Nasoz F (2004) Using noninvasive wearable computers to recognize human emotions from physiological signals. EURASIP J Appl Signal Process 2004:1672–1687
53. Lisetti C, Nasoz F, LeRouge C, Ozyer O, Alvarez K (2003) Developing multimodal intelligent affective interfaces for tele-home health care. Int J Hum-Comput Stud 59(1):245–255

54. Liu C, Conn K, Sarkar N, Stone W (2008) Physiology-based affect recognition for computer-assisted intervention of children with autism spectrum disorder. Int J Hum-Comput Stud 66(9):662–677

55. Lopes P, Liapis A, Yannakakis GN (2015) Sonancia: sonification of procedurally generated game levels. In: Proceedings of the 1st computational creativity and games workshop

56. Lopes P, Liapis A, Yannakakis GN (2015) Targeting horror via level and soundscape generation. In: Proceedings of the AAAI Artificial Intelligence for Interactive Digital Entertainment Conference

57. Malandrakis N, Potamianos A, Evangelopoulos G, Zlatintsi A (2011) A supervised approach to movie emotion tracking. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, Piscataway, pp 2376–2379

58. Mandryk RL, Atkins MS (2007) A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. Int J Hum-Comput Stud 65(4):329–347

59. Mandryk RL, Inkpen KM, Calvert TW (2006) Using psychophysiological techniques to measure user experience with entertainment technologies. Behav Inf Technol 25(2):141–158

60. Martínez HP, Yannakakis GN (2010) Genetic search feature selection for affective modeling: a case study on reported preferences. In: Proceedings of the 3rd international workshop on affective interaction in natural environments. ACM, pp 15–20

61. Martínez HP, Yannakakis GN (2011) Mining multimodal sequential patterns: a case study on affect detection. In: Proceedings of the 13th international conference on multimodal interfaces. ACM, pp 3–10

62. Martínez HP, Bengio Y, Yannakakis GN (2013) Learning deep physiological models of affect. IEEE Comput Intell Mag 9(1):20–33

63. Martinez H, Yannakakis G, Hallam J (2014) Don't classify ratings of affect; rank them!. IEEE Trans Affect Comput 5(3):314–326

64. Mateas M, Stern A (2003) Façade: an experiment in building a fully-realized interactive drama. In: Game developers conference, vol 2

65. McQuiggan S, Lee S, Lester J (2007) Early prediction of student frustration. In: Proceedings of international conference on affective computing and intelligent interaction. Springer, pp 698–709

66. Mcquiggan SW, Mott BW, Lester JC (2008) Modeling self-efficacy in intelligent tutoring systems: an inductive approach. User Model User-Adapt Interact 18(1):81–123

67. Messinger DS, Cassel TD, Acosta SI, Ambadar Z, Cohn JF (2008) Infant smiling dynamics and perceived positive emotion. J Nonverbal Behav 32(3):133–155

68. Nacke L, Lindley CA (2008) Flow and immersion in first-person shooters: measuring the player's gameplay experience. In: Proceedings of conference on future play: research, play, share. ACM, pp 81–88

69. Nagel F, Kopiez R, Grewe O, Altenmüller E (2007) Emujoy: software for continuous measurement of perceived emotions in music. Behav Res Methods 39(2):283–290

70. Nijholt A (2009) BCI for games: a 'state of the art' survey. In: Entertainment computing-ICEC 2008. Springer, pp 225–228

71. Pedersen C, Togelius J, Yannakakis GN (2010) Modeling player experience for content creation. IEEE Trans Comput Intell AI Games 2(1):54–67

72. Picard RW (2009) Future affective technology for autism and emotion communication. Philos Trans R Soc B: Biol Sci 364(1535):3575–3584

73. Picard RW, Vyzas E, Healey J (2001) Toward machine emotional intelligence: analysis of affective physiological state. IEEE Trans Pattern Anal Mach Intell 23(10):1175–1191

74. Picard RW, Papert S, Bender W, Blumberg B, Breazeal C, Cavallo D, Machover T, Resnick M, Roy D, Strohecker C (2004) Affective learning – a manifesto. BT Technol J 22(4):253–269

75. Poels K, de Kort Y, Ijsselsteijn W (2007) It is always a lot of fun!: exploring dimensions of digital game experience using focus group methodology. In: Proceedings of the 2007 conference on future play. ACM, pp 83–89

76. Poh M-Z, McDuff DJ, Picard RW (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt Express 18(10):10762–10774

77. Qu L, Wang N, Johnson W (2005) Using learner focus of attention to detect learner motivation factors. In: Proceedings of international conference on user modeling (UM). Springer, pp 149–149

78. Rani P, Sarkar N, Liu C (2005) Maintaining optimal challenge in computer games through real-time physiological feedback. In: Proceedings of the 11th international conference on human computer interaction, pp 184–192

79. Ravaja N, Saari T, Laarni J, Kallinen K, Salminen M, Holopainen J, Jarvinen A (2005) The psychophysiology of video gaming: phasic emotional responses to game events. In: Proceedings of digital games research association conference (DiGRA)

80. Ravaja N, Saari T, Salminen M, Laarni J, Kallinen K (2006) Phasic emotional reactions to video game events: a psychophysiological investigation. Media Psychol 8(4):343–367

81. Rebolledo-Mendez G, Dunwell I, Martínez-Mirón E, Vargas-Cerdán M, De Freitas S, Liarokapis F, García-Gaona A (2009) Assessing neurosky's usability to detect attention levels in an assessment exercise. In: Human-computer interaction. New trends, pp 149–158

82. Riedl M, Bulitko V (2012) Interactive narrative: a novel application of artificial intelligence for computer games. AAAI, Citeseer

83. Robison J, McQuiggan S, Lester J (2009) Evaluating the consequences of affective feedback in intelligent tutoring systems. In: Proceedings of international conference on affective computing and intelligent interaction (ACII). IEEE, pp 1–6

84. Russell JA (1980) A circumplex model of affect. J Personal Soc Psychol **39**(6):1161

85. Schwarz N (2000) Emotion, cognition, and decision making. Cogn Emot 14(4):433–440

86. Shaker N, Yannakakis GN, Togelius J (2010) Towards automatic personalized content generation for platform games. In: Proceedings of the AAAI conference on artificial intelligence and interactive digital entertainment (AIIDE). AAAI Press

87. Shaker N, Asteriadis S, Yannakakis GN, Karpouzis K (2013) Fusing visual and behavioral cues for modeling user experience in games. IEEE Trans Cybern 43(6):1519–1531

88. Sharma N, Gedeon T (2012) Objective measures, sensors and computational techniques for stress recognition and classification: a survey. Comput Methods Programs Biomed 108(3):1287–1301

89. Sokolov EN (1963) Higher nervous functions: the orienting reflex. Annu Rev Physiol 25(1):545–580

90. Tijs T, Brokken D, Ijsselsteijn W (2008) Dynamic game balancing by recognizing affect. In: Proceedings of international conference on fun and games. Springer, pp 88–93

91. Togelius J, Schmidhuber J (2008) An experiment in automatic game design. In: IEEE symposium on computational intelligence and games, CIG'08. IEEE, pp 111–118

92. Togelius J, De Nardi R, Lucas SM (2007) Towards automatic personalised content creation for racing games. In: IEEE symposium on computational intelligence and games, CIG 2007. IEEE, pp 252–259

93. Togelius J, Preuss M, Beume N, Wessing S, Hagelback J, Yannakakis GN (2010) Multiobjective exploration of the starcraft map space. In: 2010 IEEE symposium on computational intelligence and games (CIG). IEEE, pp 265–272

94. Togelius J, Yannakakis GN, Stanley KO, Browne C (2011) Search-based procedural content generation: a taxonomy and survey. IEEE Trans Comput Intell AI Games 3(3):172–186

95. Tognetti S, Garbarino M, Bonarini A, Matteucci M (2010) Modeling enjoyment preference from physiological responses in a car racing game. In: Proceedings of IEEE conference on computational intelligence and games (CIG). IEEE, pp 321–328

96. van den Hoogen WM, IJsselsteijn WA, de Kort YAW (2008) Exploring behavioral expressions of player experience in digital games. In: Proceedings of the workshop on facial and bodily expression for control and adaptation of games (ECAG), pp 11–19

97. Vogt T, André E (2005) Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. In: Proceedings of IEEE international conference on multimedia and expo (ICME). IEEE, pp 474–477

98. Yang Y-H, Chen HH (2011) Ranking-based emotion recognition for music organization and retrieval. IEEE Trans Audio Speech Lang Process 19(4):762–774

99. Yannakakis GN (2009) Preference learning for affective modeling. In: 3rd international conference on affective computing and intelligent interaction and workshops, ACII 2009, Amsterdam, Sept 2009. IEEE, pp 1–6
100. Yannakakis GN, Hallam J (2007) Towards optimizing entertainment in computer games. Appl Artif Intell 21(10):933–971
101. Yannakakis GN, Hallam J (2008) Entertainment modeling through physiology in physical play. Int J Hum-Comput Stud 66(10):741–755
102. Yannakakis GN, Hallam J (2009) Real-time game adaptation for optimizing player satisfaction. IEEE Trans Comput Intell AI Games 1(2):121–133
103. Yannakakis G, Hallam J (2011) Rating vs. preference: a comparative study of self-reporting. In: Proceedings of international conference on affective computing and intelligent interaction (ACII). Springer, pp 437–446
104. Yannakakis GN, Martínez HP (2015) Grounding truth via ordinal annotation. In: 2015 international conference on affective computing and intelligent interaction (ACII)
105. Yannakakis GN, Martínez HP (2015) Ratings are overrated! Front ICT 2:13
106. Yannakakis GN, Paiva A (2013) Emotion in games. In: Handbook on affective computing, p 20
107. Yannakakis GN, Togelius J (2011) Experience-driven procedural content generation. IEEE Trans Affect Comput 2(3):147–161
108. Yannakakis GN, Hallam J, Lund HH (2008) Entertainment capture through heart rate activity in physical interactive playgrounds. User Model User-Adapt Interact 18(1):207–243
109. Yannakakis GN, Martínez HP, Jhala A (2010) Towards affective camera control in games. User Model User-Adapt Interact 20(4):313–340
110. Yildirim S, Narayanan S, Potamianos A (2011) Detecting emotional state of a child in a conversational computer game. Comput Speech Lang 25(1):29–44
111. Zeng Z, Pantic M, Roisman G, Huang TS et al (2009) A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Trans Pattern Anal Mach Intell 31(1):39–58

Chapter 8
Emotion and Attitude Modeling for Non-player Characters

Brian Ravenet, Florian Pecune, Mathieu Chollet, and Catherine Pelachaud

Abstract Within this chapter, we are presenting how game developers could take inspiration from the research in Embodied Conversational Agent to develop non-player characters capable of expressing believable emotional and social reactions. Inspired by the social theories about human emotional and social reactions, the researchers working with Embodied Conversational Agents developed different computational models to reproduce these human mechanisms within virtual characters. We are listing some of these works, comparing the different approaches and theories considered.

Introduction

Non-player characters (NPC) that players encounter in games can have very different roles. Depending on the game itself, they can act as an obstacle, being an enemy, or they can act as an ally, helping the player to reach his/her objectives. Interacting with an NPC may be important for the progression of the player. To avoid blocking the player's progress with an unexpected situation the NPCs' behavior is usually scripted, meaning that they follow a precise predefined scenario. Therefore they usually act as emotionless robots that are only here to obey the rules of the game; they do not adapt their behavior to the current game situation, giving no sense of engagement in their interaction with the player. In order to create more compelling experiences, one can consider developing an emotional connection with the players [81]. One step toward this goal is to model NPCs with socio-emotional behaviors adapted to the game. In this chapter, we focus on the research on emotion and attitude modeling for the non-player characters in order to enhance this connection. Some games successfully convey emotional themes by proposing a very cinematographic experience and by including non-playable sequences and rich dialogues in which the characters (including the avatar of the player) can show powerful emotional behaviors like in the critically acclaimed

B. Ravenet (✉) • F. Pecune • M. Chollet • C. Pelachaud
LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France
e-mail: brian.ravenet@gaips.inesc-id.pt

video game *The Last of Us* [75]. Game with a much more narrative experience can elicit emotional connections, for example *Heavy Rain* [74] or *The Walking Dead* [77], by creating characters that express different emotional reactions during the non-playable sequences of the game depending on the choices of the player during the interactive sequences. However, these systems still use a scripted scenario and even if the developers created a very large tree of possibilities with rich emotional expressions, the NPCs' reactions during the interactive phases do not show variability depending on the social bond the player could develop with them. The video game *The Sims 4* [27] is a recent example of a game that makes use of the emotions of the virtual characters to trigger various reactions of the NPCs during the gameplay phases. However, to our knowledge, very few games endow the NPCs with enough autonomous capabilities to trigger adaptive emotional behaviors.

Autonomous virtual characters are capable of reacting in a human-like way, in any situations. In particular they are able to reason and to take decision to overcome an event [47]. When interacting with humans, they can respond emotionally and show their engagement [12], creating compelling and engaging narrative experiences. In particular, Embodied Conversational Agents [17] have been endowed with the capacity to display believable emotional and social reactions. To build such agents, computational models of emotion and social behaviors for virtual characters have been developed for more than a decade now [36]. In this chapter, we present the current state of these computational models. In the next section, we detail works on emotion modeling, first presenting different emotion theories from the literature in Human and Social sciences and then describing computational models of emotion; we also present models on emotion expressions in virtual characters. Then, in section "Attitude Modeling", we turn our attention to social behaviors like interpersonal attitudes. Finally, in section "Conclusion", we review what has been discussed and present perspectives for game developers using these technologies.

Emotion Modeling

Different computational models have been proposed in order to model triggering and expressing mechanisms linked to emotions inspired by human behaviors.

In this section, we present theories that are popular among the virtual human researchers to base their model on, and we review some existing works on emotional virtual humans.

Theory on Emotions

The literature in Human and Social sciences contains different representations for the emotions and how they are triggered. For instance, Kleinginna and Kleinginna

listed 92 different definitions that might be regrouped into 11 categories [42]. In this
section, we are presenting some of the most popular theories of emotion.

Basic Emotions

Scholars have proposed that some emotions are defined by a fixed neuromotor
program; they are biologically predetermined [26]. This theory supports Darwin's
hypothesis claiming that there is a limited number of basic emotions. These
emotions are innate; essentially they have a communicative function, and are
thus universally recognized [22]. Ekman, for instance, defines six basic emotions
corresponding to particular and distinct facial expressions: anger, disgust, fear,
sadness, happiness and surprise [25]. However, recent work tends to deny the
universality of basic emotions. In [43], authors assess that facial expressions might
be interpreted differently depending on the culture and the gaze direction of the
one expressing the emotion. By combining these basic emotions, secondary ones
might arise. In his "wheel of emotion", Plutchick compares emotions to colours
[63]. Hence, two basic emotions can be mixed together to obtain a third one; love
is considered as a mix between ecstasy and admiration. To complete the analogy,
Plutchick also considers intensity to determine emotional labels. As the colour fades
away, terror turns into fear, then apprehension.

Multidimensional Models

As opposed to the discrete approach where emotions are labeled, the continuous
approach uses a multidimensional space to represent emotions as a particular point
in this space. The main dimension used to distinguish emotions is related to *pleasure*
or *pain* [38]: indeed, this axis of valence allows one to distinguish a pleasant emotion
(e.g. joy) from a not pleasant one (e.g. sadness). However, it seems difficult to
differentiate emotions such as fear, anger or boredom by using only this single
dimension. A more accurate representation therefore requires the addition of one
or more axes on top of the valence one. In [69], Russell advocates for a model
based on two axes called "Circumplex Model of Affect" obtained by adding the
dimension of arousal to the dimension of valence. This model is taken over by
Reisenzein [67], but remains criticized, especially in [35] where the authors not
only refute the circular hypothesis, but also replace the arousal dimension by the
dimension of dominance. One of the most widely used dimensional models is the
PAD emotional model [53], which combines Pleasure, Arousal and Dominance to
obtain a better and more precise description of the different emotional states. A
recent study also confirms the universality of these three axes, while adding a fourth
dimension of unpredictability [30].

Appraisal Theory

The most recent theory in the domain of emotions is the appraisal theory, supported, for example, by Scherer [70]. According to this theory, emotions arise from a continuous evaluation of events that are happening combined with our mental state. The cognitive process can be divided into four different evaluation phases: (1) checking the relevance to know whether the event affects me or my group, (2) evaluating the impact of the event on my beliefs and goals, (3) defining my coping potential that can be used to face the situation and (4) calculating the significance of the event according to my social norms and standards. Thus, this representation is powerful enough to describe how different persons engaged in a same situation might express different emotions.

One of the most widely spread theory in the field of affective computing is the OCC theory [60] in reference to its authors Ortony, Clore and Collins. In [60] 22 emotions are categorized into 6 separate groups based on the condition that triggered them. Emotions can arise as reactions to (1) events impacting the goals of the individual, (2) events affecting standards and norms of the individual and (3) events related to the attractiveness of a particular object. The strength of this theory is that it is easily understandable and implementable. It takes into account the valence of emotions. Indeed, every emotion is considered here as pleasant or unpleasant allowing for example to avoid ambiguities such as surprise, which can be good or bad, or even neutral. In [6], the author explains how OCC can be integrated in virtual agents.

Computational Model of Emotions

According to the "Affective Loop" [76], virtual agents have to generate and express congruent emotions to allow powerful experience. In this section, we are listing some works that tried to model the complex mechanisms of emotions for virtual humans. Marsella and colleagues [50] provide an overview of the different computational models of emotion according to the theory they are based on. As stated by the authors, most of these models are rooted in the appraisal theory that was presented in the previous section. Moreover, they propose an idealized computational appraisal architecture and try to decompose the appraisal process into different modules. The appraisal derivation model transforms an event into different appraisal variables. These variables differ according to the theory of the model. Then, the affect derivation model maps the appraisal variables into a particular affective state and specifies the appropriate emotional reaction.

FAtiMA [23] follows this blueprint and offers a generic appraisal framework that can be used to compare the different appraisal theories. The framework implements a core layer (containing appraisal derivation and affect derivation models) on which additional components can be added. The cultural component, for instance, allows the agent to determine the praiseworthiness of an event according to its

cultural norms and values. The motivational component introduces basic human drives which are used to determine an event's desirability: an event raising agent's drives will be appraised as desirable while another event lowering its drives will be appraised as undesirable. EMA [49] is a computational model also based on appraisal theory. In EMA, the virtual agent interprets events using a set of appraisal variables that are stocked in a structure called *appraisal frame*. Each event is then represented by an appraisal frame, leading to a particular emotion. Every time the agent evaluates an event, the corresponding frame is updated leading to possible change in the agent's emotional state. Since many frames can be activated at the same time, the final affective state corresponds to the most recently updated frame. The model also implements two different coping strategies, altering the way the agent will evaluate new events. Some serious games for children, using the FatiMA framework, were developed to propose an emotion-based experience: *FearNot* [4] presented the danger of bullying, *ORIENt* [29] was aimed at teaching how to develop intercultural empathy and *My Dream Theater* helped children to learn how to resolve conflicts [16].

Some computational models also map the emotions calculated by their appraisal component into a three dimensional PAD (Pleasure, Arousal, Dominance) space [53]. Alma [34] is one of the first models that computes an emotion based on OCC theory and converts it into a three dimensional vector. Alma also introduces a more lasting affective state called *mood*. Once an emotion is triggered, it will *"push"* or *"pull"* the mood of the agent according to the positions of the mood and the emotion in the 3D cube. If the computed mood and the computed emotion do not belong to the same cube octant, the mood will be pulled towards the emotion. On the opposite, if the emotion is in between the current mood and the center of the cube, the mood will be pushed towards the edge of the cube. More recent works also map OCC emotions into a PAD space, like WASABI did [7] or GAMYGDALA [64]. The first one also models the mutual influence of emotions and mood over time. Emotions of positive or negative valence respectively increase or lower the mood value. Moreover, agents in a positive mood are more disposed to experience positive emotions. Following Damasio's theory [21] the model also differentiates primary emotions (represented as a point in the 3D space) and secondary emotions (represented as areas on the same space and implying a more complex cognitive process to be computed). GAMYGDALA provides a simpler generic engine that can be used to compute emotions for Non Playable Characters (NPC) in any kind of video games. To do so, game developers have only to define NPC goals, and annotate events happening in the game with a relation to these goals. According to the OCC Model, GAMYGDALA computes the related emotion for the NPC and maps it into a PAD space.

Emotions can also be represented in a formal way as a combination of logical concepts such as beliefs, desires and intentions. In [1], the authors provide a logical formalization of the emotion triggering process as described by the OCC theory using the agent's beliefs and desires. For instance, an agent experiences hope if its desire matches its expectation (i.e. the agent desires to be hired by a company and expects that it will happen). The agent will then feel satisfied, if the expected

event actually happens, or disappointed if the event doesn't happen. Meyer proposes a different formalization for four distinct emotions, namely happiness, sadness, anger and fear [54]. In this work, emotions are driven by the status of the agent's intentions. Thus, sadness is elicited if the agent believes that it cannot fulfill one of the sub-goals needed to reach its intention. The author also provides strategies to cope with the elicited emotions. However, the intensity of the different emotions is not represented by these models. An answer can be found in [59], where the authors propose different variables used to compute emotions intensity: the degree of certainty concerning the intention achievement, the effort invested to try to complete the intention, the importance of the intention and the potential to cope in case of an intention failure.

Finally, some works try to integrate a computational model of emotions into a more general cognitive architecture. Hence, emotions are part of a complex cognitive process. EMA, which has already been introduced above, has been coupled with the SOAR cognitive architecture [44]. The appraisal process presented in LIDA [32] does not really differ from the one presented in [50]. LIDA relies on Scherer's appraisal variables (relevance, implications, coping potential and normative significance) to assign an emotion to a particular appraised situation. The emotion elicited will then improve learning and facilitate later action selection, by improving the likelihood of an action to be selected. PEACTDIM [48] is another attempt to unify cognitive behavior and emotions. Based on Scherer's sequential checks, PEACTDIM adds several layers into the SOAR cognitive process. Contrary to many other computational models, emotions in PEACTDIM are not represented by a label or a multidimensional vector, but by the entire appraisal frame.

Expression of Emotions

Now that we saw how to represent and compute emotional reactions, we are going to see different works on computing the multimodal behaviors for a virtual character to display its emotional state. Defining natural and believable expressions of emotions has been one of the main topics of the ECA community since the past decade [62].

One solution is the use of motion capture performances that are reproduced, as captured, directly onto the virtual characters [28]. It is the solution adopted by most game developers. Whereas this approach has the advantage of being very realistic for a specific context, it lacks adaptability and variability. On the other hand, the approach of the ECA community is towards the creation of computational models for the real time synthesis of the emotional expressions. To build these computational models of emotional expression, researchers usually rely on two distinct approaches. The first is based on the collection of data (on human behaviors) and the identification of the features of the emotional expression within this data. The second is based on the literature in Human and Social sciences and on the findings of this literature to create rule-based systems. We now present some relevant works to illustrate the differences between these two approaches.

Data-Driven Models

Some researchers choose to use databases of expressive behaviors, from which, characteristics of the emotional behaviors can be automatically identified and extracted.

Researchers usually use motion capture to build these databases. For instance, the Emilya database [31] is a collection of motion capture data of the whole body from actors performing simple tasks in various emotional contexts. The MMLI database [56] has been built with motion capture data of people laughing in interaction. Machine learning techniques can then be applied to these databases to identify and extract features of the emotion expressions and to build computational models of emotional behaviors. For instance, in [24], the authors applied machine learning techniques on a database of people laughing. The computational model learned the relationship between body movements, facial expressions and acoustic features of laughter such as its energy and pseudo-phonemes. A different approach consists in learning directly a corpus of animation for a virtual agent without captures from human actors. For instance in [58], the authors used a crowd-sourcing method to collect a database of descriptions of different virtual agents smiles (polite, amused or embarrassed). Each description consists of values for the different parameters of the smile (e.g. degree of mouth aperture, of mouth extension). They built a decision tree, directly from this corpus, capable of choosing the values for the parameters depending on the desired type of smile.

Data driven models suffer from the need of an important amount of data. Data collection can be difficult and costly to gather but offers the advantage of obtaining an adaptable generic model that can evolve with new data.

Literature-Based Models

The literature of Human and Social sciences gives us different theories on how humans express emotions [79]. For instance, Ekman proposed a model of description of how facial expression works [26]. His system called Facial Action Coding System (FACS) is used to describe facial expression at the muscular level. FACS is often used to code facial expressions of emotions.

Some researchers, when trying to build a computational model of emotional expression, choose to derive computational rules from the findings of the literature of Human and Social sciences. Like in [78] where the authors compute the expression associated to an emotion as a linear combination of known expressions of emotions set in a 3D space. In [55], the authors present a system inspired by Scherer's appraisal theory [72] that generates sequences of multi-modal signals conveying emotions. In [45], the authors use the dimensional model of emotions Pleasure-Arousal-Dominance (PAD) to ground the different emotional contexts in which head and body movements vary during gaze shifts. Another dimensional representation (Valence-Arousal-Dominance) is used in [2] where the authors attempt to create better emotional expressions by using asymmetric facial expressions. In

[51], the authors present a system that generates the nonverbal behavior of a virtual character depending on its speech. The speech is analyzed to extract characteristics and, using rules derived from the literature, the appropriate behaviors are selected.

These systems are less costly to produce as they do not require data to power them. Moreover, the rules derived can be customized to fit particular needs (scenario, cultural or gender specific behaviors for instance) and obtain a rich repertoire of multi-modal behaviors. However they lack the adaptability and variability of the data-driven models.

Attitude Modeling

Like the modeling of emotions, modeling the attitude of virtual humans requires the development of models capable of computing and expressing attitudes.

The research about attitudes is quite new in the ECA community compared to the research about emotions. However, a few systems already exist; they rely on different theories from the literature of Human and Social sciences. In this section, we are presenting some theories about the representations and the expressions of attitudes and we are reviewing some work on social virtual humans.

Theory on Attitudes

First it is important to define what an attitude is. A review of the relevant literature is proposed in [19] where the authors present different definitions. One of them is the commonly used definition (in the ECA community) of interpersonal stances by Scherer [71]. Scherer explains that an attitude is "an affective style used naturally or strategically in an interaction with a person or a group of persons". In other words, within an interaction, one might use different attitudes depending on one's interlocutor. One might act nicely with a friend or bossy with a subordinate. These attitudes are expressed using verbal and nonverbal cues as explained in [19].

In order to replicate these attitudes within virtual humans, it is necessary to choose a representation. Different ones have been proposed through the years. The representation from Schutz consists of three dimensions which are the Inclusion, the Control and the Appreciation [73]. Later, Burgoon and Hale proposed a 12-dimensional representation to characterize different styles of interaction [13]. But the most used representation in the ECA community is the one from Argyle which consists of a dimension of Status and a dimension of Affiliation. Using these axes, an alternative circular representation has been proposed by Wiggins [80] called the Interpersonal Circumplex. These 2-dimensional representations are easy to manipulate and some researchers on human behaviors used them to describe how attitudes influence the nonverbal behavior of a person. Mehrabian, for instance, described in [52] how posture, distance and gaze can convey information about Status or Affiliation.

Computational Model of Attitudes

In this section, we are listing some works that proposed to model the complex mechanisms of attitudes within virtual humans.

As explained in [71], a social attitude is a combination of both spontaneous appraisal of the situation and strategic intentions. However, most of the computational models focus on the spontaneous appraisal, where agents only display what they feel. If an agent feels like it has power over another one, it may show dominance. Furthermore, if an agent really likes another one, it may express friendliness. Thus, to know which social attitude an agent should display toward another agent, we first have to compute its social relation.

One approach to model the dynamics of agent's social relations and thus the attitude it expresses is based on the emotions felt by the agent; that is agent displays its emotional state as a sign of its social relations toward its interlocutors. In SCREAM [66], emotions felt by the agent play an important role, changing the relationship according to their valence and intensity. A positive emotion elicited by another agent will raise the liking value towards it, while a negative emotion will have the opposite effect. The authors also add the notion of familiarity, still changing according to emotions, but evolving monotonically: only positive emotions are taken into account. Similar dynamics can be found in [57], where authors describe the influence of particular emotions on liking, dominance and solidarity. For instance, an agent A feeling an emotion of pride elicited by another agent B will improve A's values of dominance and liking toward B. These values are initially defined by the role of the agent. Finally, in EVA [40], the relation between the agent and the user is represented by two values of friendliness and of dominance. As for these works [40, 57], these values evolve according to four emotions felt by the agent: gratitude, anger, joy and distress.

In SGD [65], the authors try to team up humans with a group of synthetic characters, using a formal representation of liking and dominance. However, the evolution of these two dimensions does not rely on emotions, but on the content of the interactions between the agents. Socio-emotional actions, such as encouraging or disagreeing with one agent, will have an impact (respectively positive and negative) on its liking value. Instrumental actions, such as enhancing an agent's competence or obstructing one of its problems, will have an impact on its dominance. Callejas et al. also rely on a circumplex representation to build a computational model of social attitudes for a virtual recruiter [15]. In this work, the social attitude of the recruiter is dynamically computed according to the difficulty level of the interview and the anxiety level of the user. The recruiter will be friendly in lower difficulty levels, but might change its attitude as the difficulty increases. Here, the attitude is expressed strategically, in order to comfort or to challenge the user.

Although all the works presented above use a multidimensional representation of social attitudes, some other works only model one dimension. For example, Castelfranchi [18] formalizes the different patterns of dependence that can happen in a relationship. Basically, an agent is dependent on another one if the latter

can help the former to achieve one of its goals. The dependence level may vary if the dependent agent finds alternative solutions, or manages to induce a mutual or reciprocal dependence. Hexmoor et al. [39] address autonomy, power and dependence from another perspective. In this work, the agent's power is characterized as a difference between personal weights and liberties of preferences. The weights influence the agent towards individual or social behavior. The liberties represent internal or external processes that influence the agent's preferences of choice.

Avatar Arena focuses on the appreciation in a scenario in which a user must negotiate a schedule appointment with several agents [68]. Before each session, the appreciation level between agents is fixed (low, medium, high), as well as their assumptions about other agents preferences. According to the Congruity Theory described by Osgood and Tannenbaum [61], when an agent discovers a mismatch between its assumption about another agent's preference and what this agent's preference actually is, this might trigger a change in the appreciation level toward the other agent. Finally, some works rely on stage models to implement the notion of intimacy in their agents. This is the case for Laura, who encourages users to exercise on a daily basis [11]. Lauras behavior is driven by its intimacy level that evolves during the interactions. The more the user interacts with Laura, the more familiar it will behave. Another example of relational agent is Rea, who adapts its dialog strategy according to the principle of trust [10]. Endowing the role of an estate agent, Rea uses small-talks to enhance the confidence of the user. Once the user becomes more confident with Rea, task-oriented dialog can take place.

Expression of Attitude

Different systems that aimed at computing the behaviors expressing an attitude are presented in this section.

In the Demeanour project, virtual characters were used as avatars by users improvising a story [5]. The users can define their avatar's interpersonal attitude, and posture and gaze behavior would then be automatically generated for the avatar. For instance, a friendly avatar would orient itself more towards other avatars.

Fukayama et al. have proposed a gaze model that can express attitudes [33]. They proposed a two-state Markov model (i.e. a state where the gaze is directed at the interlocutor, a state where the gaze is averted), the parameters of which (i.e. total amount of gaze directed at the interlocutor, mean duration of gaze, direction of gaze aversion) were defined using values from the literature on gaze behavior [3]. They found that a very low (25 %) or very high (100 %) amount of gaze directed at the interlocutor conveys hostility. Dominance is linearly correlated with the amount of gaze directed at the interlocutor. Downward gaze aversions are less dominant, and sideways gaze aversions are less friendly.

The Laura ECA was developed in order to engage with users in long-term relationships [11]. The goal of Laura was to motivate the users to start a physical

activity. Two versions of Laura were compared in a longitudinal study with users: a neutral version and another version designed to appear friendlier throughout the interactions. The friendly version would produce more gestures, head movements, facial expressions of emotions (e.g. displays of empathy towards the user), and would appear physically closer to the user on the display screen. The friendly agent was attributed higher scores on a variety of measures including trust, respect, and affiliation.

Bee et al. have studied the dominance expressed by a virtual agent in a series of studies [8, 9]. In the first study [8], they investigated the impact of various expressions of emotions combined with different head positions and gaze directions on the expression of dominance. In their second study [9], Bee et al. investigated the combination of a dialogue model expressing different personalities (i.e. introvert vs extravert, agreeable vs disagreeable) and a gaze model. They found that the expressed attitude is more easily identified when the two models are used together.

Lee and Marsella have proposed a model for agents in different conversational roles, based on Argyle's attitude dimensions [46]. They collected data on behaviors of bystanders and side-participants using participants acting in an improvisation scenario where the acted characters would have different attitudes towards one another (e.g. Rio is dominant towards Harmony). Using this data, they proposed a set of rules for the behavior of these side-participants, depending on the attitudes of the characters.

Cafaro et al. proposed a model for the expression of attitudes during greetings [14]. They use previous works that studied proxemics [37] and greetings [41] to define which behaviors should display at which distances in a greeting phase so that it appears more or less friendly.

Ravenet et al. used a crowdsourcing method to build a computational model for the expression of attitudes and communicative intentions. They design an online interface where users chose the behaviors of an agent according to an instruction (e.g. "Which behaviors should the agent display to *ask* a question while appearing *friendly*?"). They collected almost 1000 answers from participants. Using these collected data, they built a Bayesian network that represents the probabilities of the occurrence of the considered behaviors depending on an attitude and a communicative intention. This network can be used to generate several combinations of non-verbal signals to communicate an intention with a given attitude, increasing the variability of behaviors of the virtual agent.

Chollet et al. proposed a behavior planning model for expressing attitude; the behavior planning model plans entire sequences of non-verbal signals instead of independent signals [20]. They call this model a Sequential Behavior Planner. The Sequential Behavior Planner takes as input an utterance to be said by the agent augmented with information on the communicative intentions and its attitude it wants to convey. This technique relies on a dataset of sequences of non-verbal signals that were annotated as carrying an attitude that were extracted from a multimodal corpus using a sequential pattern mining technique. An evaluation showed that the model manages to convey friendliness, hostility, and dominance attitudes, but that it fails to express submissiveness.

Conclusion

In this chapter, we presented various works aimed at giving virtual agents the capacity of expressing believable emotions and attitudes.

Researchers working on Embodied Conversational Agents created various computational models used to trigger in an autonomous fashion, emotional and social reactions within virtual characters. They also developed solutions for the expressions of the associated behaviors (e.g. gestures, facial expression and speech). Their models are based either on the findings of the literature on Human and Social sciences or on collected data. These researchers ran experimental studies to verify that the emotional and social behaviors exhibited by the virtual characters are understood by users and corresponded to what was expected. Nowadays, games still massively use scripted characters but as the level of realism of the narrative experiences proposed by the developers is constantly increasing, the need for a higher level of believability of their worlds is increasing too. Whereas the NPCs can show powerful emotional behaviors during cinematographic sequences, they usually lack of autonomy during interactive phases. The tools presented in this chapter can be useful for game developers in order to go a step further in creating highly immersive experiences. A player, convinced by the behavior exhibited by a NPC, who can consider it to be more than a simple robot, might think carefully about his/her actions in the game as they would impact his/her experience on an emotional level [81]. The player would be able to build his/her own experience depending on how s/he chooses to interact and bond with the NPCs. Game developers can benefit from these models and moreover they can provide the research community with valuable feedbacks on how the models perform in very rich applications.

References

1. Adam C, Herzig A, Longin D (2009) A logical formalization of the occ theory of emotions. Synthese 168(2):201–248
2. Ahn J, Gobron S, Thalmann D, Boulic R (2013) Asymmetric facial expressions: revealing richer emotions for embodied conversational agents. Comput Animat Virtual Worlds 24(6):539–551
3. Argyle M (1988) Bodily communication. University paperbacks, Methuen
4. Aylett R, Vala M, Sequeira P, Paiva A (2007) Fearnot!–an emergent narrative approach to virtual dramas for anti-bullying education. In: Virtual storytelling. Using virtual reality technologies for storytelling. Springer, Berlin, pp 202–205
5. Ballin D, Gillies M, Crabtree I (2004) A framework for interpersonal attitude and non-verbal communication in improvisational visual media production. In: Proceedings of the 1st European conference on visual media production, London
6. Bartneck C (2002) Integrating the occ model of emotions in embodied characters. In: Workshop on virtual conversational characters. Citeseer
7. Becker-Asano C (2014) Wasabi for affect simulation in human-computer interaction. In: Proceedings of international workshop on emotion representations and modelling for HCI systems

8. Bee N, Franke S, Andrea E (2009) Relations between facial display, eye gaze and head tilt: dominance perception variations of virtual agents. In: 3rd international conference on affective computing and intelligent interaction and workshops, ACII 2009, pp 1–7
9. Bee N, Pollock C, Andrea E, Walker M (2010) Bossy or wimpy: expressing social dominance by combining gaze and linguistic behaviors. In: Allbeck J, Badler N, Bickmore T, Pelachaud C, Safonova A (eds) Intelligent virtual agents. Volume 6356 of lecture notes in computer science. Springer, Berlin/Heidelberg, pp 265–271
10. Bickmore T, Cassell J (2005) Social dialongue with embodied conversational agents. In: van Kuppevelt J, Dybkjær L, Bernsen NO (eds) Advances in natural multimodal dialogue systems. Springer, Dordrecht, pp 23–54
11. Bickmore TW, Picard RW (2005) Establishing and maintaining long-term human-computer relationships. ACM Trans Comput-Hum Interact 12(2):293–327
12. Brave S, Nass C, Hutchinson K (2005) Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. Int J Hum-Comput Stud 62(2):161–178
13. Burgoon JK, Buller DB, Hale JL, de Turck MA (1984) Relational messages associated with nonverbal behaviors. Hum Commun Res 10(3):351–378
14. Cafaro A, Vilhjálmsson H, Bickmore T, Heylen D, Jóhannsdóttir K, Valgarðsson G (2012) First impressions: users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In: Nakano Y, Neff M, Paiva A, Walker M (eds) Intelligent virtual agents. Volume 7502 of lecture notes in computer science. Springer, Berlin/Heidelberg, pp 67–80
15. Callejas Z, Ravenet B, Ochs M, Pelachaud C (2014) A computational model of social attitudes for a virtual recruiter. In: Proceedings of the 2014 international conference on autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, Richland County, pp 93–100
16. Campos H, Campos J, Cabral J, Martinho C, Nielsen JH, Paiva A (2013) My dream theatre. In: Proceedings of the 2013 international conference on autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, Richland County, pp 1357–1358
17. Cassell J, Sullivan J, Prevost S, Churchill E (2000) Embodied Conversational Agents. MIT, Cambridge
18. Castelfranchi C, Miceli M, Cesta A (1992) Dependence relations among autonomous agents. Decentralized AI 3:215–227
19. Chindamo M, Allwood J, Ahlsén E (2012) Some suggestions for the study of stance in communication. In: Privacy, security, risk and trust (PASSAT), 2012 international confernece on social computing (SocialCom), pp 617–622. IEEE, Los Alamitos
20. Chollet M, Ochs M, Pelachaud C (2014) From non-verbal signals sequence mining to bayesian networks for interpersonal attitudes expression. In: Bickmore T, Marsella S, Sidner C (eds) Intelligent virtual agents. Volume 8637 of lecture notes in computer science. Springer, Cham, pp 120–133
21. Damasio A (1994) Descartes' error: emotion, reason, and the human brain. Putnam Publishing, New York
22. Darwin C (1872) The expression of the emotions in man and animals. Oxford University Press, Oxford
23. Dias J, Mascarenhas S, Paiva A (2014) Fatima modular: towards an agent architecture with a generic appraisal framework. In: Emotion modeling. Springer, Cham, pp 44–56
24. Ding Y, Prepin K, Huang J, Pelachaud C, Artières T (2014) Laughter animation synthesis. In: Proceedings of the 2014 international conference on autonomous agents and multi-agent systems, AAMAS '14, Richland. International Foundation for Autonomous Agents and Multiagent Systems, pp 773–780
25. Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. J Personal Soc Psychol 17(2):124
26. Ekman P, Friesen, WV (1978) Facial action coding system. Consulting Psychologists Press, Palo Alto

27. Electronic Arts (2014) The Sims 4, Computer Game, Maxis
28. Ennis C, Hoyet L, Egges A, and McDonnell R (2013) Emotion Capture: Emotionally Expressive Characters for Games. In *Proceedings of Motion on Games* (MIG '13). ACM, New York, pp 53–60
29. Enz S, Zoll C, Vannini N, Hall L, Paiva A, Aylett R, Lim MY (2012) Orient: the intercultural 535 empathy. In: Edmundson A (ed) Cases on Cultural Implications and Considerations in Online Learning. IGI Global, Hershey, p 282
30. Fontaine JR, Scherer KR, Roesch EB, Ellsworth PC (2007) The World of emotions is not two-dimensional. Psychol Sci 18(12):1050–1057
31. Fourati N, Pelachaud C (2014) Emilya: emotional body expression in daily actions database. In: LREC
32. Franklin S, Madl T, D'mello S, Snaider J (2014) Lida: a systems-level architecture for cognition, emotion, and learning. IEEE Trans Auton Ment Dev 6(1):19–41
33. Fukayama A, Ohno T, Mukawa N, Sawaki M, Hagita N (2002) Messages embedded in gaze of interface agents – impression management with agent's gaze. In: Proceedings of the 2002 SIGCHI conference on human factors in computing systems. ACM, New York, pp 41–48
34. Gebhard P (2005) Alma: a layered model of affect. In: Proceedings of the fourth international joint conference on autonomous agents and multiagent systems. ACM, New York, pp 29–36
35. Gehm TL, Scherer KR (1988) Factors determining the dimensions of subjective emotional space. In: Scherer KR (ed) Facets of emotion: recent research. Lawrence Erlbaum Associates, Inc, Hillsdale, pp 99–113
36. Gratch J, Marsella S (2005) Evaluating a computational model of emotion. Auton Agents Multi-Agent Syst 11(1):23–43
37. Hall ET (1969) The hidden dimension, vol 1990. Anchor Books, New York
38. Hewstone ME, Stroebe WE, Stephenson GME (1996) Introduction to social psychology: a European perspective. Blackwell Publishing, Oxford
39. Hexmoor H (2002) A model of absolute autonomy and power: toward group effects. Connect Sci 14(4):323–333
40. Kasap Z, Ben Moussa M, Chaudhuri P, Magnenat-Thalmann N (2009) Making them remember – emotional virtual characters with memory. IEEE Comput Graph Appl 29(2):20–29
41. Kendon A (1990) Conducting interaction: patterns of behavior in focused encounters, vol 7. CUP Archive, Cambridge/New York
42. Kleinginna PR Jr, Kleinginna AM (1981) A categorized list of emotion definitions, with suggestions for a consensual definition. Motiv Emot 5(4):345–379
43. Krämer K, Bente G, Kuzmanovic B, Barisic I, Pfeiffer UJ, Georgescu AL, Vogeley K (2014) Neural correlates of emotion perception depending on culture and gaze direction. Cult Brain 2(1):27–51
44. Laird J (2012) The soar cognitive architecture. MIT, Cambridge/London
45. Lance B, Marsella SC (2007). Emotionally expressive head and body movement during gaze shifts. In: Intelligent virtual agents. Springer, Berlin/New York, pp 72–85
46. Lee J, Marsella S (2011) Modeling side participants and bystanders: the importance of being a laugh track. In: Vilhjálmsson H, Kopp S, Marsella S, Thórisson K (eds) Intelligent virtual agents. Volume 6895 of lecture notes in computer science. Springer, Berlin/Heidelberg, pp 240–247
47. Luck M, Aylett R (2000) Applying artificial intelligence to virtual reality: intelligent virtual environments. Appl Artif Intell 14(1):3–32
48. Marinier RP, Laird JE, Lewis RL (2009) A computational unification of cognitive behavior and emotion. Cogn Syst Res 10(1):48–69
49. Marsella SC, Gratch J (2009) Ema: a process model of appraisal dynamics. Cogn Syst Res 10(1):70–90
50. Marsella S, Gratch J, Petta P (2010) Computational models of emotion. In: Scherer KR, Banziger T, Roesch E (eds) A blueprint for affective computing–a sourcebook and manual. Oxford University Press, Oxford, pp 21–46

51. Marsella S, Xu Y, Lhommet M, Feng A, Scherer S, Shapiro A (2013) Virtual character performance from speech. In: Proceedings of the 12th ACM SIGGRAPH/Eurographics symposium on computer animation. ACM, New York, pp 25–35
52. Mehrabian A (1969) Significance of posture and position in the communication of attitude and status relationships. Psychol Bull 71(5):359
53. Mehrabian A (1996) Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. Curr Psychol 14(4):261–292
54. Meyer J-JC (2006) Reasoning about emotional agents. Int J Intell Syst 21(6):601–619
55. Niewiadomski R, Hyniewska S, Pelachaud C (2009) Modeling emotional expressions as sequences of behaviors. In: Intelligent virtual agents. Springer, Berlin/New York, pp 316–322
56. Niewiadomski R, Mancini M, Baur T, Varni G, Griffin H, Aung M (2013) MMLI: multimodal multiperson corpus of laughter in interaction. In: Salah A, Hung H, Aran O, Gunes H (eds) Human behavior understanding. Volume 8212 of lecture notes in computer science. Springer, Cham, pp 184–195
57. Ochs M, Sabouret N, Corruble V (2009) Simulation of the dynamics of nonplayer characters' emotions and social relations in games. IEEE Trans Comput Intell AI Games 1(4):281–297
58. Ochs M, Niewiadomski R, Pelachaud C (2010) How a virtual agent should smile? In: Allbeck J, Badler N, Bickmore T, Pelachaud C, Safonova A (eds) Intelligent virtual agents. Volume 6356 of lecture notes in computer science. Springer, Berlin/Heidelberg, pp 427–440
59. Ochs M, Sadek D, Pelachaud C (2012) A formal model of emotions for an empathic rational dialog agent. Auton Agents Multi-Agent Syst 24(3):410–440
60. Ortony A, Clore GL, Collins A (1988) The cognitive structure of emotions. Cambridge 601 University Press, Cambridge
61. Osgood CE, Tannenbaum PH (1955) The principle of congruity in the prediction of attitude change. Psychol Rev 62(1):42
62. Pelachaud C (2009) Modelling multimodal expression of emotion in a virtual agent. Philos Trans R Soc B: Biol Sci 364(1535):3539–3548
63. Plutchik R (2001) The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. Am Sci 89(4):344–350
64. Popescu A, Broekens J, van Someren M (2014) Gamygdala: an emotion engine for games. IEEE Trans Affect Comput 5(1):32–44
65. Prada R, Paiva A (2009) Teaming up humans with autonomous synthetic characters. Artif Intell 173(1):80–103
66. Prendinger H, Descamps S, Ishizuka M (2002) Scripting affective communication with life-like characters in web-based interaction systems. Appl Artif Intell 16(7–8):519–553
67. Reisenzein R (1994) Pleasure-arousal theory and the intensity of emotions. J Personal Soc Psychol 67(3):525
68. Rist T, Schmitt M, Pelachaud C, Bilvi M (2003) Towards a simulation of conversations with expressive embodied speakers and listeners. In: 16th international conference on computer animation and social agents. IEEE, Los Alamitos, pp 5–10
69. Russell JA (1980) A circumplex model of affect. J Personal Soc Psychol 39(6):1161
70. Scherer KR (2001) Appraisal considered as a process of multilevel sequential checking. Apprais Processes Emot: Theory Methods Res 92:120
71. Scherer KR (2005) What are emotions? And how can they be measured? Soc Sci Inf 44(4):695–729
72. Scherer KR, Ellgring H (2007) Multimodal expression of emotion: affect programs or componential appraisal patterns? Emotion 7(1):158
73. Schutz WC (1958) FIRO: a three-dimensional theory of interpersonal behavior. Rinehart, New York
74. Sony Computer Entertainment (2010) Heavy Rain, Playstation 3, Quantic Dreams
75. Sony Computer Entertainment (2013) The Last of Us, Playstation 3, Naughty Dog
76. Sundström P (2005) Exploring the affective loop. PhD thesis, Stockholm University
77. Telltale Games (2012) The walking dead, computer, mobiles and consoles. Telltale Games

78. Tsapatsoulis N, Raouzaiou A, Kollias S, Cowie R, Douglas-Cowie E (2002) Emotion recognition and synthesis based on MPEG-4 FAPs. In: Pandzic IS, Forchheimer R (eds) MPEG-4 facial animation. Wiley, Chichester/Hoboken, pp 141–167
79. Wallbott HG (1998) Bodily expression of emotion. Eur J Soc Psychol 28(6):879–896
80. Wiggins J (1996) The five-factor model of personality: theoretical perspectives. The Guilford Press, New York
81. Yannakakis G, Karpouzis K, Paiva A, Hudlicka E (2011) Emotion in games. In: Affective computing and intelligent interaction, Memphis, pp 497–497

Chapter 9
Emotion-Driven Level Generation

Julian Togelius and Georgios N. Yannakakis

Abstract This chapter examines the relationship between emotions and level generation. Grounded in the experience-driven procedural content generation framework we focus on levels and introduce a taxonomy of approaches for emotion-driven level generation. We then review four characteristic level generators of our earlier work that exemplify each one of the approaches introduced. We conclude the chapter with our vision on the future of emotion-driven level generation.

Introduction

Game levels are frequently capable of, and indeed designed to, elicit affective responses. Such responses range from the sadness of traversing a desolate landscape, to the feeling of achievement upon clearing a hard but fair challenge, to the delight of finding a hidden treasure cache, to the frustration of butting ones head against an abusively hard challenge, to the tense dread of exploring a dark maze where a monster might appear any second. The player might experience different and changing emotions while playing a single level. The affective response of players to games are influenced by numerous factors—many of them detailed in this book, such as sound effects, narrative and cinematography—but this particular chapter will focus on level design. We will be looking at level design as the arrangement of components or items from a given vocabulary in order to yield a space for the player character(s) to progress through. This can be exemplified by the multitude of levels designed in *Super Mario Maker* (Nintendo 2015), levels which share a common and somewhat restricted set of items and affordances, but which explore a remarkably large expressive range and give rise to a wide variety of player emotions.

The structure of this paper will build on our own *experience-driven procedural content generation* framework, which describes how Procedural Content Generation

J. Togelius (✉)
Department of Computer Science and Engineering, New York University, New York, NY, USA
e-mail: julian@togelius.com

G.N. Yannakakis
Institute of Digital Games, University of Malta, Msida, Malta
e-mail: georgios.yannakakis@um.edu.mt

(PCG) methods can be used to adapt games according to models of player experience [23]. From this perspective, computer games are dynamic media that implement rich forms of user interactivity. They also allow for high levels of player incorporation and yield dynamic and complex emotion manifestations. The potential that games have to influence players is mainly due to the rich contextual building blocks (i.e., game content) they offer and their ability of placing the player in a continuous mode of interaction with the game. Players are continuously presented with (and react to) a wide palette of content types that vary from sound effects and textures to narratives, game rules and levels. This rich interactivity can naturally accommodate mechanisms for real-time adaptation of game content aimed at adjusting player experience and realizing affective interaction [23].

In the rest of this chapter we provide a brief taxonomy of approaches for emotion-driven level generation by putting an emphasis on two core dimensions that influence the relationship between level generation and affect modeling. First, we distinguish between level generation that follows a player-centric approach and that which follows a designer-centric approach. Second, we differentiate between level generation approaches that consider affect directly and approaches that are built on other aspects of player experience such as cognitive patterns and player behaviors. We provide an example for each of the four possibilities. The chapter concludes with a vision of the future of emotion-driven level generation.

Emotion-Driven Level Generation

Emotion-driven level generation can be viewed as an instance of the *experience-driven procedural content generation* framework [23]. According to our definitions in [23] *player experience* is the collection of affective patterns elicited, cognitive processes emerged and behavioral traits observed during gameplay [22]. *Game content* refers to all aspects of a game that affect the player experience but are not non-player character (NPC) behavior or the game engine itself. This definition includes game design, level architecture, visuals, audio, and narrative [9]. *Procedural content generation* (PCG) refers to the creation of game content—as defined above—automatically (or semi-automatically), through algorithmic means. As games offer one of the most representative examples of rich and diverse content creation applications and are elicitors of unique user experiences we view game content as the *building block* of games and the generated games as the *potentiators* of player experience. Based on the above, the experience-driven PCG framework [23] is defined as a generic approach for the optimization of player experience via the adaptation of the experienced content.

To realize experience-driven PCG for level generation one needs to assess the quality of the level generated (linked to the experience of the user), search through the available level content, and generate a level that optimizes the experience for the user (see Fig. 9.1). In particular, the key components of experience-driven PCG for level generation are:
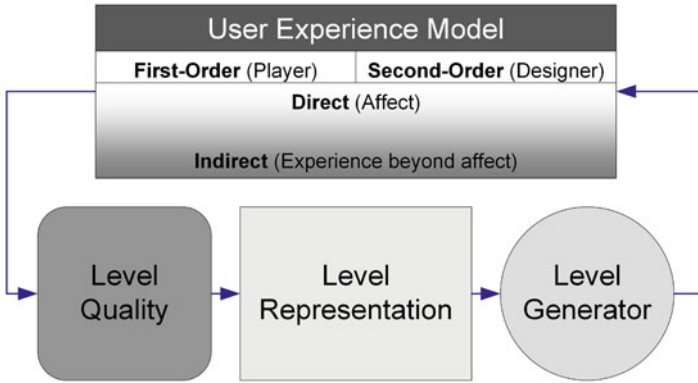
Fig. 9.1 The four key components of the experience-driven PCG framework [23] for level generation. The four user experience modeling options are detailed in the taxonomy of section "A Taxonomy of Emotion-Driven Level Generators"

- **User experience model**: user experience is modeled as a function of game content and the user. The user considered can be either a player (i.e., first-order level generation) or a designer (i.e., second-order level generation); see section "First-Order vs. Second-Order Level Generators" for further details. The modeling approach can be either direct or indirect depending on its level of grounding in user affect; see section "Direct vs. Indirect Level Generators" for further details. Section "A Taxonomy of Emotion-Driven Level Generators" introduces the taxonomy of the four aforementioned user experience modeling options for level generation, thereby, enriching the experience-driven PCG framework.
- **Level quality evaluator**: the quality of the generated content (i.e., level) is assessed and linked to the modeled experience.
- **Level representation**: the level is represented accordingly to maximize search efficacy and robustness.
- **Level generator**: the generator searches through content (i.e., parameterized level) space for content that optimizes the experience for the user according to the acquired model.

With respect to the user experience component of the experience-driven PCG framework, emotion-driven level design focuses on emotion and affect and takes into consideration other aspects of experience only implicitly (as discussed thoroughly in the following sections). With regards to the PCG component of experience-driven PCG, emotion-driven level generation considers game levels and their core architectural properties (functionality and aesthetics) as the content type under consideration. In other words emotion-driven level generation investigates the generation of game levels and their impact on gameplay and experience.

A Taxonomy of Emotion-Driven Level Generators

According to the taxonomy presented in [19, 23] game content can be *necessary* (e.g. game rules or a main quest) or *optional* (e.g. trees in a level, flying birds on the background or a side quest). Necessary content needs to be completable or playable by the player, and generators of necessary content therefore needs to assure the completeness of the generated artefacts. We here consider levels to be necessary content for a digital game as most game levels need to be completable. Further, a generator can be either *offline* or *online*, *random* or based on a *parametrizable*, *stochastic* or *deterministic* and finally it can be either *constructive* (i.e. content is generated once) or *generate-and-test* (i.e. content is generated and tested). In addition to the taxonomies provided in [19, 23]—which are applicable to level generators—in this section we put an emphasis on the level design process and derive two more dimensions for clustering level generation approaches. The two dimensions are illustrated under the *user experience model* component of Fig. 9.1.

First-Order vs. Second-Order Level Generators

Arguably the level design process as a whole is, by nature, built and driven by emotion. On the one hand there is a *player* that experiences a particular game level. That interaction with the game level elicits affective responses, enables particular cognitive processes and, as a result, yields to a particular playing behavior. Such player emotional responses may, in turn, reflect on the player's bodily reactions (facial expression, posture) or affect changes in the player's physiology. Those affect manifestations caused (in part) by the design of the level can be captured via e.g. physiological sensors or web cameras (see other chapters of this book) and can be used as input to a model that predicts player emotion. Such a model can, in turn be used for personalized level design. In this chapter, we refer to this player-centric approach to emotion-driven level generation *first-order*. On the other hand there is a level *designer* that has particular goals, intentions, preferences, styles and expectations from her design [8]. Most importantly, the level designer incrementally internalizes and builds a high level (or even rather detailed) model of expected player experience during the design process that is used as a design guide. That internal model is tested through piloting, and thorough play-testing. If via testing a mismatch is found between the model of the expected player experience and the actual player experience then two design options are applicable and can even concur: either the level is adjusted accordingly or the designer's expectations and goals about the player experience are altered to match the actual experience. The game emotive goals of the designer and aspects of that internal player experience model can be captured in a similar fashion as with the player. The designer manifests bodily, cognitive and behavioral responses to the design during the design process. Such

responses can provide the input to computational representations of the designer's affective, cognitive or behavioral aspects (i.e. *designer models* [8]). We name that designer-centric approach to player experience design as *second-order* since it is based on an indirect modeling of player experience.

In summary first-order experience-driven level generators build on a *model of player experience*, whereas second-order generators build on a *model of designer experience* which may include intents, goals, styles, preferences and expectations (see Fig. 9.1).

Direct vs. Indirect Level Generators

Further to the distinction between first- and second-order approaches to emotion-driven level generation we also identify two ways in which affect is incorporated in level generation: the *direct* and the *indirect* approach (see Fig. 9.1). According to the *direct* approach the evaluation function of the level generation mechanism is built on a computational model of the player's affect. On the other hand an *indirect* level generator instead considers other aspects of the player experience beyond affect and emotion—such as behavioral traits and cognitive processes. These aspects are seen as proxies of player experience, therefore the *indirect* label. Evidence (as well as common sense) suggests that player (or designer) actions, decisions and real-time preferences are interlinked to experience since the level may affect the player's or the designer's cognitive processing patterns and cognitive focus. As a result, cognitive processes and behavioral patterns may influence emotions and vice versa as cognition, behavior and emotion are heavily interwoven [1]. Thus, one may infer the player's or the designer's emotional state indirectly by analyzing patterns of the interaction and associating user emotions with level context variables [2, 3]. Given the interwoven relation of affect and cognition the boundaries that distinguish between a direct and an indirect level generation approach are often unclear. Figure 9.1 depicts this relationship via a gradient-colored pattern.

Exemplifying Emotion-Driven Level Generation

In this section we outline one critical emotion-driven level generation example per each of the four categories derived from the taxonomy presented in section "A Taxonomy of Emotion-Driven Level Generators". Within the first-order level generation approaches we describe the *Super Mario Bros* (direct) and the *Mini Dungeons* (indirect) paradigms whereas within the second-order level generators we present the *Sonancia* (direct) and the *Sentient Sketchbook* (indirect) tools for level design.

Super Mario Bros: First-Order, Direct Level Generation

Building on the experience-driven PCG [23] framework, the Super Mario Bros level generator employs a *direct* and *first-order* approach to emotion-driven level design. The work of Pedersen et al. [14, 15] and Shaker et al. [16–18] focuses on the construction of models of player affect via crowdsourced gameplay traces and self-reports of the player experience of several hundred Super Mario Bros players. The resulting models fuse behavioral characteristics of gameplay with level parameters and predict aspects of player experience such as player challenge, frustration and engagement. These modes can, in turn, be used to generate personalized emotion-driven levels by varying the level parameters considered by the player experience models.

More specifically, the work of Shaker et al. [18]—which builds upon and extends that of Pedersen et al. [14, 15]—mines a large set of crowdsourced gameplay data of Super Mario Bros. The data consists of 40 short game levels that differ along six key level design parameters. Collectively, these levels are played 1560 times over the Internet and the perceived experience is annotated by participants via self-reported rankings of engagement, frustration and challenge. The study explores dissimilar types of features, including direct measurements of event and item frequencies, and features constructed through frequent sequence mining. The fusion of the extracted features allowed Shaker et al. to predict reported player experience with accuracies higher than 70 %. The models of engagement, frustration and challenge contain level parameters as their input and, thus, are directly applicable for the personalization of game experience via automatic level generation. Exhaustive search within the level parameter space has been used in [17] to achieve that aim.

In addition to the large data corpus of behavioral cues, level parameters and subjective experience annotations a sequel article of Shaker et al. [16] investigated the impact of player *visual cues* (obtained via a webcam) for the construction of player experience models. Obtained results show that when players' visual and behavioral characteristics are fused highly accurate experience models can be constructed as accuracies reach 91 %, 92 %, and 88 % for engagement, frustration, and challenge, respectively. Using exhaustive search on the small level parameter space models can be used to generate a sample of maximally (or minimally) engaging, frustrating, and challenging levels (see Fig. 9.2a).

MiniDungeons: First-Order, Indirect Level Generation

MiniDungeons is a simple turn-based dungeon crawling game, in the style of popular roguelikes such as *Desktop Dungeons* (with similarities to games such as *Rogue* and *NetHack* [4, 5]). The gameplay consists in navigating maze-like dungeons to get from the entrance to the exit of each dungeon (see Fig. 9.2b). Typically, there is more than one way of reaching the end and multiple dead ends. Scattered around

Fig. 9.2 The example level generators discussed in this chapter. Super Mario Bros is a trademark of Nintendo; all images used with permission. (**a**) A *Super Mario Bros* level that maximizes the frustration of a particular player [16]. (**b**) *Mini Dungeons*: a screenshot from a generated level. (**c**) *Sonancia*: a screenshot from a generated level. (**d**) The *Sentient Sketchbook* strategy map design tool

the dungeon are monsters, treasures and potions. Monsters sometimes block the path to the exit and need to be overcome to win the level, other times they block paths to treasures or just stand around in the open. Fighting monsters drains health, which can be regained by consuming potions. Importantly, the game can be played in many different ways, depending on whether the player focuses on finishing levels quickly, getting all the treasures, killing all the monsters etc., and also depending on how risk-averse the player is.

Holmgård et al. developed a method for modeling players' behavior in the MiniDungeons game (and, by extension, other games featuring tactical decisions) from the perspective of bounded rationality. The model assumes a small number of

objectives and takes parameters specifying how important each objective is to the player. Using neuroevolution, agents can be trained to replicate a player's style—at least those aspects of player style which are captured by a set of objective weights.

Being able to replicate a player's playing style is very useful for level generation. Liapis et al. designed a level generator for MiniDungeons based on simulated playthrough [6]. The generator uses evolutionary search for in level space, using playthroughs of levels for evaluating them. By feeding the generator a specific player model, the generator will create levels tailored to the modeled playstyle in the sense that playstyle is very successful at that level. This is a *first-order* and *indirect* level generator, because while it models the player, it does not model player experience directly; instead, it models the player's playstyle. It is assumed that the player wants to play in the particular style they exhibit, and therefore that generating levels that make that playstyle successful will increase player enjoyment.

Sonancia: Second-Order, Direct Level Generation

Sonancia [11, 12] is a system built for generating multiple creative domains of horror games, with the intention of creating tense and frightful experiences. *Sonancia* procedurally generates the architecture of a haunted mansion (with rooms and doors which may contain monsters or quest items) as well as the level's soundscape by allocating audio assets within the rooms and mixing them as the player traverses the level (see Fig. 9.2c). Level generation and soundscape generation are orchestrated by notions of tension and suspense; the level generator attempts to match a designer-specified progression of tension while the sound generator attempts to prompt the player's suspense in rooms where tension is low.

The *Sonancia* level and soundscape generation system is *direct* as it relies on a function that maps sound and level features to a tension model. The tension-driven level generation is also *second-order* as it explicitly depends on a designer's provided tension curve—which, in turn, implies the existence of an indirect model of player experience. Further details about the current level and sound generation algorithm behind *Sonancia* can be found in [11, 12].

Sentient Sketchbook: Second-Order, Indirect Level Generation

The last remaining quadrant of our taxonomy is occupied by the *second-order*, *indirect* level generators. These are generators that model the designer, but not the designer's affective experience directly. In the following, we will discuss the example of *Sentient Sketchbook* with its designer modeling component.

Sentient Sketchbook is an AI-assisted game design tool for strategy game maps, such as those used in *StarCraft* (Blizzard Entertainment 1998) [7]. At the core, there is a standard level editor featuring abilities to sketch a strategy map (see

Fig. 9.2d). The tool constantly measures the qualities of the current state of the level design through several metrics related to exploration, area control and balance, and provides real-time feedback to the designer as well as suggestions for changes that the designer can choose to apply or ignore. In the graphical user interface, the various metrics are visualized as meters that give the user instant feedback about e.g how resource-balanced the current version of the level is, but there is also a visualization in the actual editor pane for e.g. safe resources. The suggestions are presented in a separate panel to the right of the editor, and the user can at any time choose to use a suggestion. These suggestions are generated partly using evolutionary algorithms, starting from the current level and trying to find level variants that satisfy some of the level metrics better.

The designer modeling in Sentient Sketchbook [10] works by constantly tracking the quality metrics of the level as it is being edited. The model then tries to estimate the trend in the editing; essentially, estimate the gradient in multidimensional quality space. This model is then used to influence what suggestions are generated. In a nutshell, the suggestions are generated to mostly lie in the direction the user seems to be pursuing. So if a designer using Sentient Sketchbook seems to be aiming for a more asymmetric map where player A has the resources and player B has the more advantageous terrain, most of the suggestions the tool presents will follow that trend and be even more asymmetric in terms of resources and terrain.

In sum, Sentient Sketchbook with its designer modeling component implements second-order indirect level generation, in that it models the designer's intent and acts on this model. The emotional expression of the levels and the elicited player experience are assumed to be implicit in the intent of the designer, and the model is helping the designer to implement this intent through the generation of suggestions.

Discussion

While the examples discussed here come from academic research, it is worth noting that *dynamic difficulty adjustment* is a widespread practice within commercial games of some genres. In particular racing games (such as *Mario Kart 64* (Nintendo 1996)) frequently adapt their difficulty based on the performance of the player. Some other commercial games include more complex mechanisms; in particular *Left 4 dead* (Valve 2008) uses a sophisticated dynamic difficulty adjustment mechanism based on tension curves. Player experience, however, is a more complex synthesis of affective and cognitive patterns than mere challenge and only certain aspects of it have been modeled in games. Explicit player emotion-based adaptation exists in commercial games such as the biofeedback-based game *Journey of Wild Divine* (Wild Divine 2001) for relaxation purposes or the adventure horror biofeedback-enhanced game *Nevermind* (Flying Mollusk 2015). A number of sensors are available for affective interaction with those games including skin conductance and heart activity. Nevertheless the emotion-based adaptation is realized either through audiovisual aspects or the challenge offered to the player. At the time of writing, we

are not aware of any commercial games that explicitly model player experience for the purpose of generating levels.

In order for emotion-driven game adaptation through level generation to become widespread in commercial-standard games, a number of questions need to be answered, presumably through further research. These questions deal with what features are effective for modeling player experience, how best to generate levels given a particular experience model, and the stability and generality of acquired models. Another critical question is how often particular level attributes should be adjusted. The frequency can vary from simple predetermined or dynamic time windows [21] but adaptation can also be activated every time a new level [17] or a new game [20] starts, or even after a set of critical player actions—such as in *Façade* [13]. The time window of adaptation is heavily dependent on the game under examination and the desires of the game designer.

Future Vision and Conclusion

We have outlined a taxonomy of approaches to emotion-driven level generation, elaborating on our existing taxonomy of experience-driven procedural content generation. We have also discussed four examples of emotion-driven level generation, one for each corner in our two-dimensional taxonomy. This is existing work, but it would be safe to say that only very little of the potential of emotion-driven level generation has been realized at this point in time. As so often, further work is needed. Therefore, we would like to conclude this chapter with a brief vision of what a game might look like in the future, when we have figured out emotion-driven level generation sufficiently to make it work reliably on a large-scale commercial-grade game.

You are playing an "open world" game, something like *Grand Theft Auto V* (Rockstar Games 2013) or *Skyrim* (Bethesda Softworks 2011). Instead of going straight to the next mission objective in the city you are in, you decide to drive (or ride) 5 h in some randomly chosen direction. The game makes up the landscape as you go along, and you end up in a new city that no human player has visited before. In this city, you can enter any house (though you might have to pick a few locks), talk to everyone you meet, and involve yourself in a completely new set of intrigues and carry out new missions. If you would have gone in a different direction, you would have reached a different city with different architecture, different people and different missions. Or a huge forest with realistic animals and eremites, or a secret research lab, or whatever the game engine comes up with.

While creating those areas, the game takes your skills, preferences and emotional state into consideration. All of those have been estimated earlier on through recording your interactions with the game, using models of player affect inferred from a large number of players' interactions in multiple games. So the game might infer that you are bored with the current selection of assassination quests, and venture an educated guess that some opportunities for (in-game) romance

might spice things up. Or decide that you need more, or less, challenge. Or that your aesthetic sense might be stirred by some grand open vistas accompanied by a bombastic score, or may a dark claustrophobic basement accompanied by a minimalist electronic tune. Maybe you need more content and activities similar to what you have already experienced; perhaps you had a tough day in the real world and want the comfortable embrace of well-known (yet superficially new) in-game territory and tasks.

Doing all of this right will require enormously wide-ranging and accurate models. How far can we realistically get towards this goal given current technologies and paradigms? We don't know. All we know is that more research is needed. While the methods we have today can already be implemented in constrained domains and controlled environments (e.g., see the Super Mario Bros experiments discussed above) there is no shortage of further research towards making the goal of emotionally adaptive games a reality. In other words, you (and we) have a lot to work on.

References

1. Bechara A, Damasio AR (2005) The somatic marker hypothesis: a neural theory of economic decision. Games Econ Behav 52(2):336–372
2. Conati C (2002) Probabilistic assessment of user's emotions in educational games. Appl Artif Intell 16(7–8):555–575
3. Gratch J, Marsella S (2005) Evaluating a computational model of emotion. Auton Agents Multi-Agent Syst 11(1):23–43
4. Holmgård C, Liapis A, Togelius J, Yannakakis GN (2014) Evolving personas for player decision modeling. In: Proceedings of the IEEE conference on computational intelligence and games (CIG)
5. Holmgård C, Liapis A, Togelius J, Yannakakis GN (2014) Personas versus clones for player decision modeling. In: Proceedings of the international conference on entertainment computing (ICEC)
6. Liapis A, Holmgård C, Yannakakis GN, Togelius J (2015) Procedural personas as critics for dungeon generation. In: Applications of evolutionary computation. Springer, Cham, pp 331–343
7. Liapis A, Yannakakis G, Togelius J (2013) Sentient Sketchbook: computer-aided game level authoring. In: Proceedings of the ACM conference on foundations of digital games
8. Liapis A, Yannakakis GN, Togelius J (2013) Designer modeling for personalized game content creation tools. In: Proceedings of the AIIDE workshop on artificial intelligence & game aesthetics
9. Liapis A, Yannakakis GN, Togelius J (2014) Computational game creativity. In: Proceedings of the fifth international conference on computational creativity, pp 285–292
10. Liapis A, Yannakakis GN, Togelius J (2014) Designer modeling for sentient sketchbook. In: 2014 IEEE conference on computational intelligence and games (CIG). IEEE, pp 1–8
11. Lopes P, Liapis A, Yannakakis GN (2015) Sonancia: sonification of procedurally generated game levels. In: Proceedings of the ICCC workshop on computational creativity & games

12. Lopes P, Liapis A, Yannakakis GN (2015) Targeting horror via level and soundscape generation. In: Proceedings of the AAAI artificial intelligence for interactive digital entertainment conference
13. Mateas M, Stern A (2005) Procedural authorship: a case-study of the interactive drama façade. Digital Arts and Culture
14. Pedersen C, Togelius J, Yannakakis GN (2009) Modeling player experience in super mario bros. In: IEEE symposium on computational intelligence and games, CIG 2009. IEEE, pp 132–139
15. Pedersen C, Togelius J, Yannakakis GN (2010) Modeling player experience for content creation. IEEE Trans Comput Intell AI Games 2(1):54–67
16. Shaker N, Asteriadis S, Yannakakis GN, Karpouzis K (2013) Fusing visual and behavioral cues for modeling user experience in games. IEEE Trans Cybern 43(6):1519–1531
17. Shaker N, Yannakakis GN, Togelius J (2010) Towards automatic personalized content generation for platform games. In: Proceedings of the AAAI conference on artificial intelligence and interactive digital entertainment (AIIDE). AAAI
18. Shaker N, Yannakakis GN, Togelius J (2013) Crowdsourcing the aesthetics of platform games. IEEE Trans Comput Intell AI Games 5(3):276–290
19. Togelius J, Yannakakis GN, Stanley KO, Browne C (2011) Search-based procedural content generation: a taxonomy and survey. IEEE Trans Comput Intell AI Games 3(3):172–186
20. Yannakakis GN, Hallam J (2007) Towards optimizing entertainment in computer games. Appl Artif Intell 21(10):933–971
21. Yannakakis GN, Hallam J (2009) Real-time game adaptation for optimizing player satisfaction. IEEE Trans Comput Intell AI Games 1(2):121–133
22. Yannakakis GN, Paiva A (2013) Emotion in games. In: Handbook on affective computing, p 20. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199942237.001.0001/oxfordhb-9780199942237
23. Yannakakis GN, Togelius J (2011) Experience-driven procedural content generation. IEEE Trans Affect Comput 2(3):147–161

Chapter 10
Emotion-Driven Narrative Generation

Brian O'Neill and Mark Riedl

Abstract While a number of systems have been developed that can generate stories, the challenge of generating stories that elicit emotions from human audiences remains an open problem. With the development of models of emotion, it would be possible to use these models as means of evaluating stories for their emotional content. In this chapter, we discuss Dramatis, a model of suspense. This model measures the level of suspense in a story by attempting to determine the best method for the protagonist to avoid a negative outcome. We discuss the possibilities for Dramatis and other emotion models for improving intelligent generation of narratives.

Introduction

Games are one of several common forms of entertainment that makes use of narrative. Many game genres use fictional context to reinforce the immersion within the game world and to motivate the player's activities. These fictional contexts answer the question, "Why am I, the player, engaging in a particular activity?" The fictional context may further induce an affective response from a player: dramatic tension over how events are unfolding, strong positive or negative feelings towards virtual characters, or suspense over what might happen next.

In many game narratives, the narrative was pre-determined by the authors and designers. The player has little or no capacity for affecting the events of the story, because the game has a single, linear narrative arc. In contrast, interactive narrative is a form of digital interactive experience in which users create or influence a dramatic storyline through actions, either by assuming the role of a character in

---

B. O'Neill (✉)
Department of Computer Science and Information Technology, Western New England University, 1215 Wilbraham Road, Springfield, MA 01119, USA
e-mail: brian.oneill@wne.edu

M. Riedl
School of Interactive Computing, Georgia Institute of Technology, 85 Fifth Street, Atlanta, GA 30308, USA
e-mail: riedl@cc.gatech.edu

a fictional virtual world, issuing commands to computer-controlled characters, or directly manipulating the fictional world state. The simplest form of an interactive narrative is a branching story, such as Choose-Your-Own-Adventure books and hypertexts, in which plot points are followed by a number of options that lead to different, alternative narratives unfolding. More complex interactive narrative systems use artificial intelligence (AI) to construct the story on the fly in accordance with the player's desires. In AI-driven interactive narrative, a *drama manager*—an omniscient background agent that monitors the state of the fictional world— conducts a search through possible future narrative trajectories, determines what will happen next in the game, and coordinates virtual characters to bring about the best narrative possible. The difficulty lies within recognizing which of the trajectories is most interesting to the player.

Good narratives are not simply a series of events—good narratives elicit emotional responses from their audiences. However, generating stories that intentionally elicit an emotional response is challenging. Maintaining an emotional level in an interactive narrative requires interference from an intelligent manager, keeping the story on trajectories that are expected to keep emotional content high. Models of emotion could help address these issues. With an understanding of emotional responses to stories, these models could be developed and used to generate emotion-inducing stories. In this chapter, we describe Dramatis, a model of suspense. With such a model, we can judge the level of suspense a reader or player would feel from following the narrative of a particular trajectory. We will also discuss the consequences for having such a model (or models of other emotions) for story generation and interactive narrative.

Background

Before discussing the Dramatis system or how we can judge the suspense level of a particular future trajectory, we must clarify what we mean by narrative and suspense. We will also briefly discuss story generation and interactive narrative, while providing examples of such systems that have attempted to address suspense and related emotional responses.

Narrative

Narrative is ubiquitous in human culture. Narratives are used in a variety of forms of entertainment, including books, films, and games. In addition to entertainment, people create and share narratives in order to explain the world around them. Prince defines narrative as follows [25]:

> **Narrative:** The representation …of one or more real or fictive events communicated by one, two, or several (more or less overt) narrators to one, two, or several (more or less overt) naratees.

Put another way, a narrative is the communication of events (rather than simple facts) from a narrator to a reader or listener. The key to the definition is the requirement for events. A fact (e.g., "It is snowing.") does not constitute a narrative. By contrast, a single event ("I went to the store.") is a narrative, albeit not one that is particularly interesting. The "main incidents" of a narrative come together to form a plot. Plots may follow a common structure, such as the traditional Aristotelian arc [2], or Freytag's triangle [11].

Narratologists distinguish between the events of the story (known as the *fabula*) and the presentation of those events by the narrator(s) to the narratee(s) (known as the *sjuzhet*). The *fabula* contains all events of the narrative, regardless of the order of presentation, or whether they are presented to the audience at all. A *sjuzhet* is a particular ordering and presentation of a subset of the events contained in the fabula. It is common for the sjuzhet to exclude events from the fabula, or to alter the order of events. A fabula may therefore have multiple sjuzhets, depending on what the narrator chooses to include from the set of events and the order in which they are told. As an example of fabula and sjuzhet, consider the original *Back to the Future* film. The events of the film, including those not shown on-screen, make up the fabula. The presentation shown to the audience is one sjuzhet. An alternate sjuzhet would show the events in chronological order.

Story Generation and Interactive Narrative

Story generation refers to the ability of artificial intelligence to create new stories. Computational approaches to story generation largely take one of two approaches: search-based approaches [17, 23, 27, 31] or adaptive approaches [13, 22, 30]. Search-based systems explore a space of possible sequences of actions, using some heuristic of quality to compare them. Adaptive systems start from a library of known stories. These stories are modified or recombined into new stories, sometimes using analogical reasoning. In some cases, story generation systems work to create both a new fabula or sjuzhet. However, some systems focus on only creating one or the other (e.g. generating a sjuzhet for a given fabula).

Interactive narrative is a form of story generation that features the audience as a user who can influence the narrative as it progresses [26]. The user, typically in the role of the protagonist, can take actions in the story-world, thereby affecting the path and outcome of the story. Some interactive narratives give the user control over the world as an observer, rather than giving them direct control of a character. Non-player characters (NPCs) in the interactive narrative may

be controlled by an experience manager, which affects the world in order to maintain story quality. The definition of quality varies depending on the particular system, but may include closeness to the author's intended story or measures of player experience, such as the expected emotional impact on the player.

A number of story generation and interactive narrative systems have attempted to address audience emotion in some way. Suspenser [7] is a story generation system which, given a fabula, attempts to identify the most suspenseful sjuzhet. Similarly, Prevoyant [3] uses flashbacks and foreshadowing to reorder a story with the goal of creating the most surprising sjuzhet from a fabula. Ware et al. [31] developed a model of narrative conflict using planning, applicable to story generation and interactive narrative. Façade [16] and Merchant of Venice [24] are interactive narrative systems which establish ideal tension curves for the narrative. The drama manager in each system affects the story by trying to get the tension in the story to the pre-defined level. Other story generation and interactive narrative systems [4, 22] also use tension as a metric, though there is no consistent definition of tension among them.

Suspense

Expert storytellers who craft their narratives for entertainment often structure their sjuzhets with the intent of eliciting emotional responses from the narratees (readers, game players, film viewers, etc.). The idea that story structure is correlated with audience enjoyment dates to Aristotle [2]. Suspense is one of many commonly used tools for creating emotional responses and has been found to contribute to reader enjoyment [29]. There are many definitions of suspense, coming from the fields of narratology [1, 6, 29], psychology [8, 12, 21], and entertainment theory [33], to name a few. Rather than consider each of these definitions, we will highlight the similarities in those definitions.

There are four attributes that are common among the various definitions of suspense: (1) uncertainty about an outcome, (2) a particularly desirable or undesirable possible result to that uncertain outcome, (3) an audience affinity for the character whose outcome is uncertain, and (4) a disparity of knowledge between the characters and the audience. The uncertainty of an outcome is the most important feature, and there must be meaning behind this uncertainty. There must be a substantial possibility of an undesirable state resulting for the character. However, there is no suspense unless the audience cares about the character. If the audience does not like the character or cannot identify with the character, then they will not feel suspense about the character's outcome. Finally, suspense can be generated by giving the audience more knowledge about a situation than the characters have. In such cases, the audience will be aware of the potential dire consequences of a situation, while the characters may have no idea.

One definition of suspense that we wish to highlight comes from Gerrig and Bernardo [12]

> Readers feel suspense when led to believe that the quantity or quality of paths through the hero's problem space has become diminished.

Gerrig and Bernardo generated this definition by studying the levels of suspense self-reported by readers given different versions of story excerpts. Readers act as problem-solvers on behalf of the protagonist, attempting to identify solutions that avert a negative outcome for the protagonist. When readers struggle to find solutions, or only find low-quality solutions, readers perceive more suspense. Thus, in a sense, readers find themselves in an interactive narrative in their own mind, as they evaluate potential future trajectories for the protagonist. However, unlike in an interactive narrative, the reader lacks the ability to decide for the protagonist which path to take.

Dramatis adopts a reformulation of Gerrig and Bernardo's definition. This reformulation is discussed in the "Reformulating Gerrig and Bernardo" section below. Suspenser [7] applies Gerrig and Bernardo's definition of suspense in its attempt to find the most suspenseful sjuzhet for a given fabula. The system measures suspense by projecting all possible future plans and determining the ratio of failed plans to successful plans, with suspense increasing as the ratio increases.

Dramatis

Dramatis is a computational model of suspense felt by the reader of a story. The model reads a story and calculates the level of suspense over time. The model uses a reformulation of Gerrig and Bernardo's definition of suspense (details of which are discussed in the next section). Dramatis reads a discretized symbolic-logic version of story events, determines whether characters are facing an undesirable outcome, and generates and evaluates the quality of the best plan for avoiding that outcome. The evaluation of quality is correlated with the level of suspense at that moment of the story.

Reformulating Gerrig and Bernardo

Recall Gerrig and Bernardo's definition of suspense, introduced above: "Readers feel suspense when led to believe that the quantity or quality of paths through the hero's problem space has become diminished." Gerrig and Bernardo describe a search space, where the search is conducted by the reader on behalf of the hero of the story. The search space consists of possible future states of the story world. Readers, therefore, are searching through a series of potential storylines and judging which one is best. Suspense is generated, in part, by how authors manipulate the space.

How do authors induce suspense? According to Gerrig and Bernardo's definition and their studies of readers, authors can manipulate the quantity and/or quality of paths in the hero's space. Authors may propose possible solutions, potentially implying an increase in quantity or that the plan is high quality, before striking it down, thereby diminishing the quantity of plans available. Authors may otherwise indicate suggest plans that they know to be faulty in order to distract readers from the solution that will ultimately be used by the hero. So authors create suspense by manipulating the search space or how the reader traverses the search space.

Gerrig and Bernardo's definition of suspense is computationally intractable. It is not possible to measure all paths through the problem space in terms of success and failure and weigh the ratio, because paths may fail as a result of the search process or the planning problem, rather than as a result of the conditions of the story world. Additionally, Gerrig and Bernardo's definition suggests that humans regenerate the search space repeatedly while reading. However, many of the definitions of suspense indicated that humans only search the space when prompted to do so by a potential undesirable outcome. Further, regenerating the search space requires the ability to identify the causal consequences of story events, an inference that can only occur when the reader puts the story aside [14]. Finally, human memory is resource-bounded, and they are therefore incapable of considering the entire space of possible events, let alone constantly regenerating that search space as they read.

As a consequence, we reformulate Gerrig and Bernardo's definition of suspense as follows:

> Given the belief that a character can face a negative outcome, one can assume that this outcome will occur and search for the single most likely plan—the *escape plan*—in which the protagonist avoids this outcome.

Gerrig and Bernardo refer to the quality of paths through the hero's problem space, though they are not precise with how quality is measured. We consider the escape plan's quality to be its perceived likelihood of success from the perspective of the reader. Using perceived likelihood (rather than actual likelihood) allows us to account for ways in which the author might manipulate the problem space, as well as account for the disparity in knowledge between the characters and the audience. We use a model of reader memory to calculate the perceived likelihood, working from the concept that humans consider the first thought retrieved from memory to be the most likely thing to actually occur [15]. Additionally, this reformulation requires neither constant regeneration of the search space, nor generation of the total search space. Finally, by searching for a single escape plan, there is no comparison between the number of successful plans and the number of failed plans.

Dramatis Algorithm and Inputs

Figure 10.1 shows the Dramatis algorithm for measuring suspense in a story. Subsequent sections break down these steps in further detail. Dramatis reads stories in a discretized symbolic format, which we call *time-slices*. Each time-slice describes one action in the story and provides state information about the characters and location of the scene. As Dramatis reads the story, it searches a library of scripts to identify one whose sequence of events matches the events observed in the story-so-far. The script provides information about what negative outcomes may occur in the near future. Once a negative outcome has been predicted, Dramatis generates an escape plan to avert the negative outcome. The perceived likelihood of the escape plan's success is correlated with the level of suspense at that moment. As the story continues and new information is gained in the story, Dramatis revises its escape plans and its measures of likelihood and suspense, potentially generating new plans as old escape plans cease to be viable. At the conclusion of the story, we are left with a curve showing suspense over time.

Let SL be a script library, OL be an operator library, and T be an ordered sequence of time-slices. Let NM represent narrative memory, MS be a model of reader salience, and EP be an escape plan.

```
DRAMATIS(SL, OL, T):
    NM ← ∅
    MS ← ∅
    EP ← ∅
    EscapePlans[] ← ∅
    Suspense[] ← ∅
    For each time-slice t ∈ T:
        Read t into NM
        Scr ← Retrieve-Script(NM, SL)
        MS ← Update-Salience-Model(NM, Scr, MS)
        If t is the next action in EP:
            cost ← Recalculate-Plan-Cost(EP, MS)
        Else:
            Links[] ← Identify-Links-To-Break(Scr, NM)
            For each link L ∈ Links:
                ⟨plans[L], costs[L]⟩ ← Generate-Escape-Plan(L, MS, NM, OL)
            cost ← min(costs[L])
            Let EP be the plan in plans[] with minimum cost in costs[]
        EscapePlans[t] ← EP
        Suspense[t] ← cost
    Return ⟨EscapePlans, Suspense⟩
```

Fig. 10.1 Dramatis algorithm

```
Action: deliver-food (Waitress, vodkaMartini, James_Bond)
Location: Casino Royale
Characters: Waitress, James_Bond
Effects: at(James_Bond pokerTable) place(pokerTable)
```

Fig. 10.2 Example time-slice

Time-Slices

As input, Dramatis requires a story, which it receives as an ordered set of discretized time-slices. Each time-slice describes exactly one action in the story. Time-Slices contain a representation of the event in the form of an instantiated STRIPS operator, the characters in the scene, the location of the scene, and the effects of the action that occurred. After the time-slice is read, its contents are stored in narrative memory, where Dramatis tracks the state of the story-world as it continues reading. In some cases, a time-slice may also contain a reference to an opposing character's plan. These plans are used to infer possible negative outcomes that the protagonist may face.

Figure 10.2 shows an example time-slice. In this time-slice, the STRIPS operator is named deliver-food, with the parameters Waitress, vodkaMartini, and James_Bond. The location of the time-slice is Casino Royale, while Waitress and James_Bond are annotated as characters. The effects listed will be added to narrative memory, along with the expected effects of the deliver-food operator. This time-slice does not contain a reference to an opposing character's plan.

Scripts

In addition to the story, Dramatis receives a script library as input. In general, scripts are conceptual frameworks used by people to navigate everyday situations (e.g., ordering food from a restaurant) [28]. When we go to restaurants, we are familiar with the typical pattern of being seated, ordering drinks, ordering food, eating the food, getting the check, and paying the check. However, when someone tells a story about eating at a restaurant and leaves out of those steps from their story, we are capable of inferring that it occurred. For example, when hearing a story about someone eating steak at a restaurant, we are capable of inferring that the steak had previously been ordered. This capacity for inference comes from scripts. Scripts, therefore, are inherently useful in understanding stories [9].

Scripts are typically learned through personal experience. However, people may also learn scripts through second-hand experiences or by hearing stories. For example, one might learn, through reading several fairy tales, that princes rescue damsels-in-distress. Thus, one could develop a script in which if a damsel is in distress, then she may be saved by a prince.

Traditionally, scripts are linear structures [28]. The scripts used by Dramatis are more akin to graphs, containing multiple possible paths that a story could take. Nodes in the graph represent events, while edges may represent temporal links or causal links. A temporal link is a directed edge in the graph, indicating that the source node occurs before the destination node in the graph. A causal link, also a directed edge, indicates that the source node provides one of the causal conditions necessary for the event in the destination node. The link is further annotated with what that condition is. Dramatis will use these scripts to infer negative outcomes (see "Predicting Negative Outcomes"). Nodes containing negative outcomes are annotated as such. Additionally, the scripts will be used in order to determine the goal situation for generating an escape plan (see "Generating Escape Plans").

Planning Operators

Finally, Dramatis is given a set of STRIPS planning operators [10] as input. STRIPS operators are described by an action, as well as the actors and objects needed to complete them. The operators define what conditions must be true in the world before the action can occur and what new conditions are true once the action has been completed. We use STRIPS operators for two purposes. First, each node in the script library represents an action and can therefore be bound to a STRIPS operator. Second, we use STRIPS operators to represent the actions that are available to characters when developing an escape plan.

Predicting Negative Outcomes

Dramatis reads a given story one time-slice at a time. As Dramatis reads, it attempts to predict whether the reader should expect a negative outcome for the protagonist. To make this prediction, Dramatis attempts to match the observed sequence of events (collected from the time-slices) to one of the scripts in its library. When choosing between scripts, Dramatis prefers scripts that make use of recently-observed story events and contains actions matching those observed earlier in the story. As the story continues, Dramatis tracks the script, maintaining a reference to the event in the script that was most recently observed in the story. Dramatis identifies the potential failure by traversing the script graph until a node is reached that is labeled as a negative outcome. Dramatis can conduct a similar process that makes use of knowledge of the antagonist's plans (when such plans have been relayed to the audience) rather than using a script. In such instances, the antagonist's plan is treated as though it were a script.

Measuring Reader Salience

After reading a time-slice, Dramatis adds each story element in the time-slice (characters, objects, locations, and events) to a model of reader memory. Dramatis uses the *Modified Event Indexing with Prediction* (MEI-P) model, which is based on previous psychological theories of the mental models of readers. Zwaan et al. [34] developed the Event Indexing (EI) model in order to understand how readers' conceptualizations of a story changed while they read. As part of his INFER system, Niehaus created the Modified Event Indexing (MEI) model [18] to account for narrative focus and readers' ability to draw inferences about the story while reading. MEI-P extends the MEI model by representing possible future events which are extracted from scripts.

MEI-P is a spreading activation network, where greater activation of a story element implies that the element is more salient in reader memory. Thus, a story element that has greater activation is more easily retrieved from memory. This ease of retrieval will be a factor in calculating the cost of actions when generating escape plans.

After Dramatis reads a time-slice, it creates a new node in the network which represents the new event. This node is connected to nodes representing the other story elements in the time-slice, creating new nodes if necessary. Each new edge in the network is given a weight of 1.0, while older connections in the network see their weights decay over time. Node activations are recalculated after each time-slice by giving them an initial activation of 1.0 and iteratively spreading node weights according to edge weights until activation levels throughout the network stabilize.

MEI-P includes predicted future events as well as observed events. Predicted events come from the script containing the negative outcome. Any event in the script that may follow from the most recently observed event is included in the MEI-P network. The salience of a predicted event decreases with how far in the future they are expected to be from the most recent of the story, just as older events decay with distance from the current event.

Generating Escape Plans

Let us review why Dramatis generates escape plans. The reformulation of Gerrig and Bernardo's definition of suspense states that when we can anticipate a negative outcome for the protagonist, the level of suspense will be correlated with the likelihood of the best plan that the reader (or Dramatis, simulating the role of the reader) can generate to avert the negative outcome. We determine the likelihood of a plan by determining the salience of story elements involved in the plan.

In order to generate an escape plan, we must first define the planning problem. The initial state of the planning problem is defined as the current state of the story world, constructed according to the information conveyed in time-slices. The goal

situation is determined by identifying the causal links in the active script between the most recently observed event and the negative outcome. If any of those links were cut, then a causally necessary condition for the negative outcome would no longer be true in the story world. As a consequence, it would no longer be possible for the negative outcome to occur, at least on the current path through the script. Thus, the goal situation is determined by negating the causal conditions in the script. Thus, an escape plan is any series of actions that leads to a violation of one or more necessary conditions for the negative outcome, thereby averting that outcome.

Each STRIPS operator's cost is calculated using the level of activation of the corresponding nodes in the MEI-P salience model. This includes the activation of the characters and objects used in the operator, any locations referenced by the operator, the preconditions and effects of the event represented by the operator, and the activation of the event itself if it is part of the model. Recall that we assume that the plan that is most easily retrieved from memory will be perceived as the most likely plan to succeed [15]. Therefore, the plan that uses elements that are most easily retrieved will be considered most likely to succeed. As a result, the cost of an operator is inversely related to the activation of the story elements used by the operator. The cost of the plan is equal to the sum of the action costs. The level of suspense is equal to the total cost of the plan. Dramatis uses the Heuristic Search Planner (HSP) [5] to generate escape plans, though any planner that can return a near-optimal result for operators with non-uniform costs would suffice.

As Dramatis continues to read the story, it tracks the most recently generated escape plan. When newly observed events conform to the events predicted by the escape plan, then we recalculate the suspense level based on the remainder of the plan and the updated MEI-P salience model. Otherwise, Dramatis generates a new escape plan using the procedure described above. It is possible that Dramatis generates the same escape plan, even though the newest event was not part of the predicted escape plan. Thus, Dramatis generates an escape plan, and therefore a suspense rating, after each time-slice of the story. We use these suspense ratings to generate a *suspense curve*, showing the change in suspense over time. Dramatis was evaluated by comparing the generated suspense curves to ratings produced by humans reading text versions of the same stories [19, 20]. Figure 10.3 shows a suspense curve created by Dramatis during this evaluation.

Future of Emotion-Driven Story Generation

Emotional models, such as Dramatis, provide an opportunity to increase the emotional content of artificially generated stories and interactive narratives. Why should these narratives be emotion-driven? Emotional content is more interesting and more entertaining [29]. Without emotion, we have a sequence of boring events. Emotional content, including suspense, makes stories worth hearing, reading, and playing. Dramatis provides a model of suspense. With other models of emotion, it will be possible to intelligently author narratives that can induce the entire spectrum of human emotions from their audiences. Dramatis, and other models of
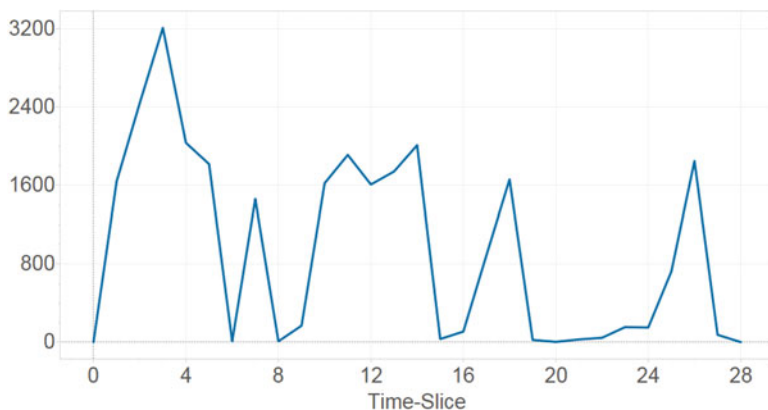
Fig. 10.3 A sample suspense curve created by Dramatis

emotion, constitute a means of direct, theory-driven evaluation of the quality of story content, under the Experience-Driven Procedural Content Generation (EDPCG) framework [32]. Under that same framework, Dramatis is an example of a model-based player—or perhaps in this case, audience—experience model. Dramatis does not address other aspects of the framework—most notably, content generation. However, Dramatis, or any other narrative emotion model, can easily serve as heuristic or evaluation function for a story generation or interactive narrative system.

Additionally, a story generation or interactive narrative system could leverage Dramatis by reasoning about the escape plans generated by the model. Suppose Dramatis were given an incomplete story to evaluate. It would read the story, generate an escape plan for the last known event of the story, and generate a suspense rating. The story generation system would then consider what events to add to the story-in-progress that would increase or decrease the level of suspense. Suspense is increased by inserting events that reduce the viability and perceived likelihood of the escape plan, while decreasing suspense requires making the escape plan seem more likely to succeed. In the case of story generation, it may be possible to insert or remove events from the middle of the incomplete story to alter the suspense level. (This would be less valuable for interactive narrative, as it would require going back in the story after the player has already progressed.) This process could continue iteratively, adding events and regenerating escape plans, until a minimal threshold of suspense is crossed, or until a pre-defined ideal suspense curve is matched. Future emotional models may have elements (akin to Dramatis escape plans) that allow for similar iterative processes.

Concluding Remarks

Stories that elicit emotion from their audiences are better, more entertaining stories. The ability to generate emotional stories automatically leads to the capacity for good

stories-on-demand. This ability bodes well for games. When games can be created with emotional stories, both human-authored and intelligently generated, then it will be possible for games to be tailored to an individual player's emotions. A game can be made with an unlimited supply of stories, where each story affects the player's emotions differently, and without the need to author the stories well in advance.

References

1. Abbott HP (2008) The Cambridge introduction to narrative. Cambridge University Press, Cambridge
2. Aristotle (1992) The poetics. Prometheus Books, Buffalo
3. Bae B, Young RM (2008) A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach. In: Proceedings of the 2008 international conference on interactive digital storytelling. Springer, Heidelberg, pp 156–167
4. Barber H, Kudenko D (2008) Generation of dilemma-based interactive narratives with a changeable story goal. In: Proceedings of the 2nd international conference on intelligent technologies for interactive entertainment. ICST, Brussels, pp 1–10
5. Bonet B, Geffner H (2001) Planning as heuristic search. Artif Intell 129:5–33
6. Branigan E (1992) Narrative comprehension and film. Routledge, New York
7. Cheong Y (2007) A computational model of narrative generation for suspense. PhD dissertation, North Carolina State University, Raleigh
8. Comisky P, Bryant J (1982) Factors involved in generating suspense. Hum Commun Res 9:49–58
9. Cullingford RE (1981) SAM and micro SAM. In: Schank R, Riesbeck CK (eds) Inside computer understanding. Lawrence Erlbaum Associates, Hillsdale
10. Fikes RE, Nilsson NJ (1971) STRIPS: a new approach to the application of theorem proving to problem solving. Artif Intell 2:189–208
11. Freytag G (1968) The technique of the drama: an exposition of dramatic composition and art. Johnston Reprint Corporation
12. Gerrig RJ, Bernardo ABI (1994) Readers as problem-solvers in the experience of suspense. Poetics 22:459–472
13. Gervás P, Díaz-Agudo B, Peinado F, Hervás R (2005) Story plot generation based on CBR. Knowl-Based Syst 18:235–242
14. Graesser AC, Singer M, Trabasso T (1994) Constructing inferences during narrative text comprehension. Psychol Rev 101:371–395
15. MacLeod C, Campbell L (1992) Memory accessibility and probability judgements: an experimental evaluation of the availability heuristic. J Personal Soc Psychol 63:890–902
16. Mateas M (2002) Interactive drama, art and artificial intelligence. PhD dissertation, Carnegie Mellon University, Pittsburgh
17. Meehan J (1981) TALE-SPIN. In: Schank RC, Riesbeck CK (eds) Inside computer understanding. Lawrence Erlbaum Associates, Hillsdale
18. Niehaus J (2009) Cognitive models of discourse comprehension for narrative generation. PhD dissertation, North Carolina State University, Raleigh
19. O'Neill B (2013) A computational model of suspense for the augmentation of intelligent story generation. PhD dissertation, Georgia Institute of Technology, Atlanta
20. O'Neill B, Riedl M (2014) Dramatis: a computational model of suspense. In: Proceedings of the twenty-eighth AAAI conference on artificial intelligence (AAAI-14). AAAI, Menlo Park, pp 944–950
21. Ortony A, Clore GL, Collins A (1988) The cognitive structure of emotions. Cambridge University Press, Cambridge

22. Pérez y Pérez R, Sharples M (2001) MEXICA: a computational model of a cognitive account of creative writing. J Exp Theor Artif Intell 13:119–139
23. Porteous J, Cavazza M (2009) Controlling narrative generation with planning trajectories. In: Proceedings of the 2nd international conference on interactive digital storytelling. Springer, Heidelberg, pp 234–245
24. Porteous J, Teutenberg J, Pizzi D, Cavazza M (2011) Visual programming of plan dynamics using constraints and landmarks. In: Proceedings of the 21st international conference on automated planning and scheduling. AAAI, Menlo Park, pp 186–193
25. Prince G (2003) Dictionary of narratology. University of Nebraska Press, Lincoln
26. Riedl MO, Bulitko V (2013) Interactive narrative: an intelligent systems approach. AI Mag 34:67–77
27. Riedl MO, Young RM (2010) Narrative planning: balancing plot and character. J Artif Intell Res 39:217–268
28. Schank RC, Abelson RP (1977) Scripts, plans, goals, and understanding: an inquiry into human knowledge structures. Lawrence Erlbaum Associates, Hillsdale
29. Tan ES (1996) Emotion and the structure of narrative film: film as an emotion machine. Routledge, New York
30. Turner SR (1993) Minstrel: a computer model of creativity and storytelling. PhD dissertation, University of California, Los Angeles
31. Ware SG, Young RM (2014) Glaive: a state-space narrative planner supporting intentionality and conflict. In: Proceedings of the 10th AAAI conference on artificial intelligence and interactive digital entertainment. AAAI, Menlo Park
32. Yannakakis GN, Togelius J (2011) Experience-driven procedural content generation. IEEE Trans Affect Comput 2:147–161
33. Zillman D (1996) The psychology of suspense in dramatic exposition. In: Vorderer P, Wulff HJ, Friedrichsen M (eds) Suspense: conceptualizations, theoretical analyses, and empirical explorations. Lawrence Erlbaum Associates, Mahwah, pp 199–231
34. Zwaan RA, Langston MC, Graesser AC (1995) The construction of situation models in narrative comprehension: an event-indexing model. Psychol Sci 6:292–297

Chapter 11
Game Cinematography: From Camera Control to Player Emotions

Paolo Burelli

Abstract Building on the definition of cinematography (Soanes and Stevenson, Oxford dictionary of English. Oxford University Press, Oxford/New York, 2005), game cinematography can be defined as the art of visualizing the content of a computer game. The relationship between game cinematography and its traditional counterpart is extremely tight as, in both cases, the aim of cinematography is to control the viewer's perspective and affect his or her perception of the events represented. However, game events are not necessarily pre-scripted and player interaction has a major role on the quality of a game experience; therefore, the role of the camera and the challenges connected to it are different in game cinematography as the virtual camera has to both dynamically react to unexpected events to correctly convey the game story and take into consideration player actions and desires to support her interaction with the virtual world. This chapter provides an overview of the evolution of the research in virtual and game cinematography, ranging from its early focus on how to control and animate the virtual camera to support interaction to its relationship with player experience and emotions. Furthermore, we will show and discuss a number of emerging research directions.

Introduction

Cinematography, over the last two centuries, has undergone a constant evolution: from the first experiments with machines such as the zoetrope [28] to the last advancements in three-dimensional cinematography and computer graphics. Throughout its history, it developed into a complex field dealing with techniques and methods to present the visual discourse. With the advent of three-dimensional computer graphics, a new branch of cinematography developed called virtual cinematography. At first limited to special effects and short films, after the first release of a fully computer generated film [43], virtual cinematography has gradually become a field on its own, even if largely intertwined with its traditional counterpart.

P. Burelli (✉)
Aalborg University, A.C. Meyers Vænge 15, 2450 København, Denmark
e-mail: pabu@create.aau.dk

One of the most clear distinctions between the two fields lies in the differences between a virtual and a real-world camera. A virtual camera is an abstract construct that defines the way the virtual world is presented to the user; it is designed to simulate the behaviour of a real-world camera and, at the current state-of-the-art in computer graphics, images produced through a virtual camera can be indistinguishable real footage. However, contrary to a real-world camera, a virtual camera has no physical presence in the virtual world and its properties can change over time and adapt to the events filmed [30].

These advantages allow virtual cameras to film scenes with much higher freedom and expressiveness; however, this can potentially contrast with the filming conventions developed in traditional cinematography. Such conventions describe aspects like the way the camera should be placed or the way it should be moved to make a transition between two different scenes, and adherence to these conventions is often important to generate a *cinematographic* experience and not to disorient the viewer – e.g. by crossing the line of action [1].

Furthermore, new application areas such as interactive narratives and computer games have further expanded the difference between real and virtual cinematography as new conventions and methods needed to be studied to address the specific characteristics of these new media. Interactive narratives, for instance, reduce the control of the designer over the *mise en scene* – i.e. the arragement of the elements in the scene – thus requiring more flexible and intelligent methods to control the camera.

Researchers have been studying effective and efficient solutions to address these differences and assist virtual designers and programmers to translate cinematographic conventions to virtual cinematography. Early research works focused on the problem of manually handling a virtual camera through input devices such as a keyboard and a mouse [66]; thereafter, the focus gradually shifted towards the problem of automating and assisting camera movements so that the camera could be animated in complex and constantly changing environments [23]. This is the case, for instance, of interactive narratives, in which the events of a story depend on the choices and actions of the viewer. This medium, which can be in many ways seen as natural evolution of films, is one of the dominant application areas in the current state-of-the-art, and different researchers have investigated the translations of classic cinematography conventions to this medium: [38], for example, have investigated the automatic generation of shot plans for emergent stories, while [45] have investigated automatic camera placement and animation in such contexts.

Beyond the differences in terms of unpredictability of the environment and events, the purpose of virtual cinematography in interactive narratives is, in large part, similar to non-interactive narratives: the camera has to support the storytelling. However, if interactive narrative is analysed as a component of a medium such as computer games, the focus of the cinematography and the purpose of the camera shifts away from solely supporting narration.

Computer games are a highly interactive medium and the virtual camera is responsible for supporting both the interaction and the visualisation of the game events, and, while traditionally the virtual camera supports narration and interaction

in different phases of a game [30], there are a number of commercial examples of games in which these two aspects overlap [18, 49, 54, 65]. Furthermore, different studies on the relationship between player emotions and cognition reveal that, in computer games, virtual cinematography is deeply intertwined with player experience. Evidence suggests that this relationship goes beyond the conventions of classical cinematography [12, 17, 46].

Both the role of the player as well as its relationship with the virtual camera and the visual experience define a clear distinction between classical cinematography and *game cinematography*. In game cinematography, the player has an active role in changing the game events thus it directly influences the movements of the camera. Furthermore, to affect the player experience, in game cinematography, the camera needs to be aware of the current state of the player both inside and outside of the game, thus, creating an indirect relationship between the player and the camera.

In this chapter, we present the concept of game cinematography from its foundations in virtual cinematography to the latest studies on player experience. We start by giving an overview of the game industry's perspective on game cinematography in section "Camera Control in Computer Games". In section "Automatic Camera Control", we present the state of the art in camera animation and placement. In section "Story-Driven Interactive Cinematography", we discuss different methods to create plans of camera movements and shots to present an interactive story. In sections "Camera and Player Interaction" and "Affective Cameras", the focus shifts on the relationship between the player and different aspects of game cinematography. Finally, in section "Future Directions in Game Cinematography", we highlight a number of possible future directions of research in game cinematography.

Camera Control in Computer Games

Game cinematography has shown a low degree of experimentation in the game industry, especially in comparison to other aspects such as rendering or physics. With few exceptions, there is a strict dichotomy between interactive cameras and cinematic cameras [30] in which the first one is used during the gameplay, while the second one is used for storytelling during cut-scenes.

Christie et al. [23] divide the camera control styles in games in the following three categories:

First person: The camera position and orientation corresponds to the player's location in the virtual environment; therefore, the camera control scheme follows the character control scheme. Examples of games adopting such a camera control scheme include *Doom* [36] and *Halo: Combat Evolved* [11].

Third person: The camera shows the events in the game from an external perspective. This perspective can be freely controllable by the player or bound to specific locations, orientations or characters. Examples of games using such type of camera control paradigm are action games such as *Tomb Raider* [24],

Fig. 11.1 Examples of advanced camera control in modern computer games. (**a**) A screen-shot from Heavy Rain [18], demonstrating usage of cinematographic techniques. Heavy Rain is a trademark of Sony Computer Entertainment (Image used with permission). (**b**) A screen-shot from Gears of War [6] during a running action; in such context the camera moves downward and shakes to enhance the haste sensation. Gears of War is a trademark of Epic Games (Image used with permission)

in which the camera follows the character from a fixed distance with different angles to avoid obstacles in the environment, or strategy and sport games – e.g. *Starcraft* [8] – in which the camera is freely movable by the player who can select different targets. In another form of third-person camera control scheme, which [30] calls *pre-determined*, multiple cameras are pre placed around the environment and, during the game, the perspective switches between them – e.g. Devil May Cry [40].

Cut-scenes and replays: In these non-interactive phases of the games, the camera focuses on representing the important elements of the story without the need to support interaction. It is often used in sport games (replays) and in story-heavy games (cut-scenes). Games featuring such a camera control scheme include *Metal Gear Solid* [41] or most sport video games.

In recent years, the separation between interactive and cinematic cameras is becoming less distinct as more games are employing cinematographic techniques to portrait narrative and interaction in games. Examples such as *Heavy Rain* [18] or *Silent Hill* [65] show extensive usage of cinematic techniques to frame the in-game actions (see Fig. 11.1a). In such games, however, the cameras are set manually in place during the development of the game; reducing heavily the movement and the actions the player can take. Furthermore, achieving the same level of quality in a game in which the content is not known in advance (e.g. it is procedurally generated [56, 68] or crowd sourced, such as in *World Of Warcraft* [9]) is still an open challenge.

Some custom dynamic techniques have been implemented in different games to achieve a more cinematographic experience in more action oriented games. For instance, in *Gears Of War* [6] (see Fig. 11.1b), the camera changes relative position and look-at direction automatically to enhance some actions or to allow for a better view of the environment. Another example is the slow motion feature implemented in the Max Payne series [55]. One of the few examples of a general camera control

system capable of handling aspects such as composition and camera movements in dynamic environments has been developed by 10Tacle Studios [32] as an extension of a system proposed by [10].

Automatic Camera Control

Under different labels, the research in virtual cinematography has distributed along three main directions: planning and definition of the shots, automatic composition and real-time camera animation. The first can be described as the task of defining a sequence of shots to visualise one or more events in a virtual environment. The second is the process of translating these shots in actual camera configurations, while the last one is the process of animating the camera during the shots and ensuring smoothness between them.

One of the first examples of a system addressing automatic camera placement and animation was presented by [7] who designed a system to automatically generate views of planets in a space simulator of NASA. Although limited in its expressiveness and flexibility, Blinn's work inspired many other researchers trying to investigate efficient methods and more flexible mathematical models to handle more complex aspects such as camera motion and frame composition [1].

More generic approaches model camera control as a constraint satisfaction problem. These approaches require the designer to define a set of desired frame properties, which are then modelled either as an objective function to be maximised by the solver or as a set of constraints that the camera configuration must satisfy. These constraints describe how the frame should look like in terms of object size, visibility and positioning. Bares et al. [3] presented a detailed definition of these constraints, which became the standard input of most automatic camera control methods. Examples of a few of these constraints can be seen in Fig. 11.2: fore example, in Fig. 11.2a, the projection size of the character is set to 37 % of the frame area, while in Fig. 11.2b, the occluded area of the rightmost character is set to a maximum of 10 % of its overall projected area.
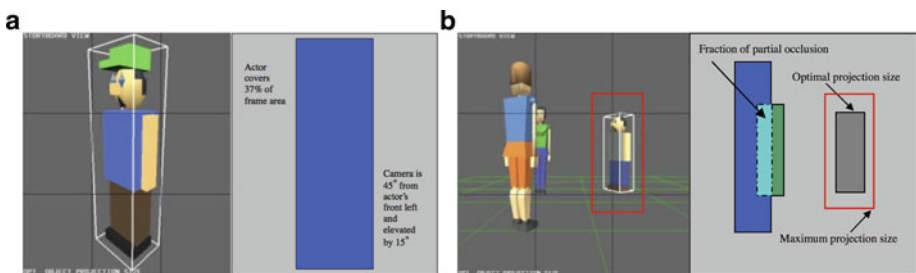
Fig. 11.2 Examples of frame constraints and their relative geometrical definition for a medium three-quarter shot and three character shot (Image used with permission and adapted from [3])

The problem of finding one or more camera configurations that satisfy a given set of frame constraints has been initially tackled by [4] using a constraint satisfaction method with a constraint relaxation technique to identify and unselect incompatible constraints. This approach used a bi-dimensional spherical map, which proved inaccurate when multiple far subjects were evaluated at the same time. Bares et al. [3] extended this initial approach by defining a sub-space of valid camera configurations for each constraint; the intersection of the valid spaces is then sampled to find the best camera configuration. The same principle of combining constraint satisfaction and search to find the best camera configuration has been further extended by improving the volume selection and integrating more sophisticated search algorithms [13, 45].

Pure optimization approaches, such as CAMPLAN [50] or the Smart Viewpoint Computation Library [53], implement a more flexible search strategy that models all frame constraints as an objective function (a weighted sum of each constraint) allowing for partial satisfaction of any constraint. These approaches do not prune any part of the search space and the satisfaction of the different frame constraints is prioritized by associating a weight to each constraint. The flexibility of such approaches comes with the price of a high computational cost. This aspect becomes a particularly critical factor when the algorithm is intended to deal with real-time dynamic virtual environments. In this context, the controller has to be able to calculate a reasonable camera configuration at short intervals (a few milliseconds) to be able to ensure synchronization with the scene changes and to have minimal impact on the overall application execution.

A more efficient approach to optimization for camera composition consists of employing local-search to find the best solution. Beckhaus et al. [5] investigated first the application of local search algorithms to camera control. Their system used Artificial Potential Fields (APFs) to guide the camera through a museum and generate smooth virtual tours. Bourne et al. [10] proposed a system that employed sliding octrees to guide the camera to the optimal camera configuration. Burelli and Jhala [14] extended these two approaches to include frame composition and support multiple-object visibility. Local search approaches offer reasonable real-time performance as they perform a small sampling of all the possible camera configurations; however, they are often unable to calculate correct camera configurations when visibility for a specific subject is required on the frame [15].

Ensuring visibility of one or more subjects is one of the most critical objectives of virtual camera composition as object visibility plays a key role in frame composition [23]. Evaluation of a subjects' occlusion can be either integrated in the search process as one part of the objective function to be optimized or as an extra heuristic to prune the search space or guide the search process. Bourne et al. [10] propose to override the current search process in case of missing visibility of the tracked subject by introducing a cut; their approach, however, considers just one object of interest. Pickering [52] suggested a shadow-volume based approach to identify volumes of space without occlusion: the search process would take place only in these volumes. Christie et al. [22], in a similar fashion, prune the space of the possible camera configurations by calculating a visibility volume – i.e. a space

of camera configurations in which visibility is guaranteed – for one or more targets; moreover, this approach takes into consideration also the temporal aspect of the occlusion to make the camera more robust to temporary and sudden occlusions.

Story-Driven Interactive Cinematography

The research works described in the previous section focus primarily on the translation of high-level cinematographic requirements (e.g. an object's visibility or its position on the screen) into low level camera parameters (e.g. position and rotation). A number of researchers have instead focused on the problem of identifying the best high-level requirements for a specific event in a virtual environment with the objective of creating a coherent a cinematographically correct visualisation of a given story. Figure 11.3 shows and example of a list of camera actions described as abstract shots (third column on the left side), which should be used to drive the camera and visualised the story described in the plan on the right side of the picture.

The first research work focusing on shot planning was published by [21]: they proposed a language (DCCL) to define shot sequences and to automatically relate such sequences to events in the virtual world. Each shot is encoded in an idiom that describes also the conditions in which the shot should be selected. He et al. [34] extended the concept of idioms within DCCL by modelling them as hierarchical finite state machines and allowing for richer expressiveness.

McDermott et al. [47] developed further this idea by allowing conditional transitions between idioms and a visual definition of the shot for each idiom. Likewise, [19] expanded the logic employed to select a shot by proposing a number of semantic rules to prioritise the shooting of specific objects or actions during the story. El-Nasr [27] followed a slightly different direction in her work, employing reactive planning and focusing on the integration of the shot selection process and scene lighting.

Based on the same principle – i.e. visualising a story from a given set of events – different researchers have refined further the aforementioned planning approaches [39] and have shown different applications such as comics generation [61] and game replays generation [25].

One common aspect among these approaches is their focus on story-visualisation: on one hand only a handful of these studies explicitly target non-interactive productions [25, 61], on the other hand, the studies that target interactive narratives, focus primarily on the emergency aspect of the narrative rather than the user's interaction. In other words, user interaction is seen just as the cause of the changes in the narrative that in turn drive the changes in the cinematography, which aims at correctly supporting the communication of the new narrative.
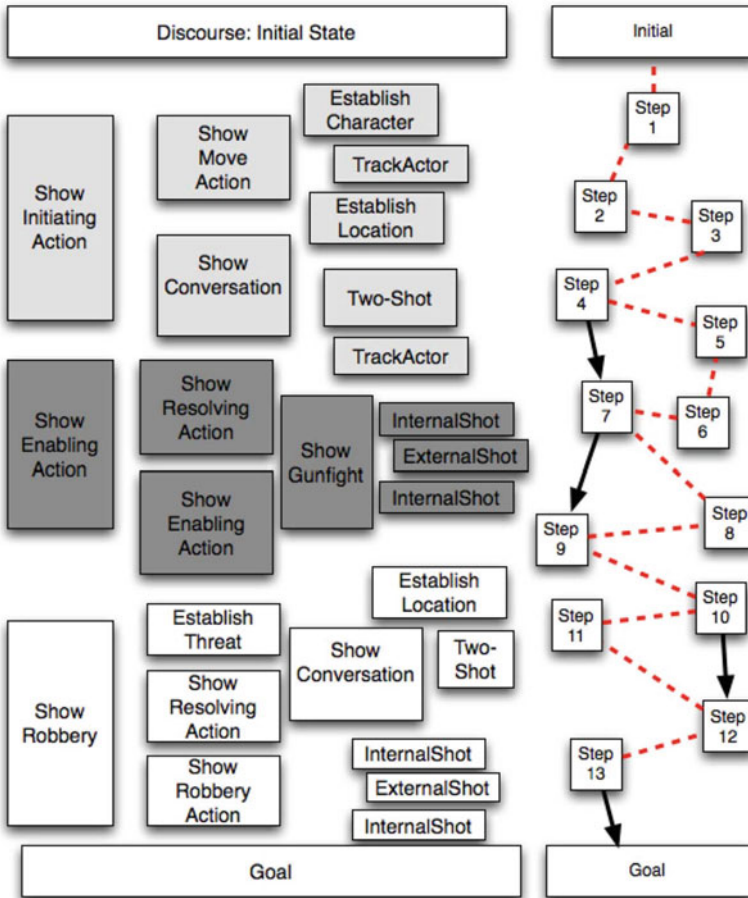
Fig. 11.3 An example of a cinematic discourse plan generated (*left side*) from a story plan (*right side*) (Image used with permission and adapted from [39])

Camera and Player Interaction

Although, story-driven cinematography can be easily identified as the dominant approach to virtual and game cinematography, a number of alternative approaches taking into consideration player preferences and player interaction have been proposed.

One of the first studies following this direction has been presented by [4]. In their work, they investigated how the user can influence the cinematographic experience and proposed a system that selects the most appropriate camera settings depending on the user's tasks in a virtual learning environment. Halper et al. [31] followed

a similar direction by adjusting the camera to accommodate the player's actions; furthermore, they devised a mechanism to predict short-term player movements to improve the camera animation smoothness.

Both the work by [31] and the one by [4] adapt the camera requirements based on a number of predefined directives designed to support user interaction; however, while the behaviour of the system would adapt depending on the task, the user preferences and the effect on the user experience is not taken into consideration. Bares and Lester [2] investigated the idea of modelling the camera behaviour according to the user preferences to generate a personalised cinematographic experience. In the system they proposed and evaluated, the user model construction required the user to specifically express some preferences about the style for the virtual camera movements.

The results of the evaluation of the work by [2] highlighted, for the first time, the importance of the relationship between the viewpoint and user interaction and suggested a direction to follow to leverage this relationship. However, the profile building procedure suggested was explicit and required the users to be conscious about the desired camera behaviour and also to be sufficiently competent to instruct the system to achieve such a behaviour.

One method to extrapolate information about the above aspects has been employed by [60] who conducted an experimental study to analyse players' gaze behaviour during a maze puzzle solving game. The results of their experiment show that gaze movements, such as fixations, are heavily influenced by the game task. They conclude that the direct use of eye tracking during the design phase of a game can be extremely valuable to understand where players focus their attention, in relation to the goal of the game.

Picardi et al. [51] and Burelli and Yannakakis [16] investigated the possibility to employ players' gaze to build user models of camera behaviour; in these works, the virtual camera behaviour is modelled on the amount of time the player spends framing and observing different objects in a virtual environment while playing a game. Combining camera movements with eye movements (i.e. fixations and pursuits) in a visually rich virtual environment such as a computer game, allows to identify exactly which objects drive the player attention [37, 62] and, therefore, can be used to build a user model of visual attention.

This model has been later employed by [17] who extended the idea of user modelling of camera behaviour [2] by implicitly building the models from in-game player behaviour and its relationship with the player's eye movements. As show in Fig. 11.4, through these models, the camera controller detects in real-time what objects will the player desire to see and it can generate appropriate camera requirements to keep these objects on screen. The results of the study by [17] show that adapting the camera behaviour based on user models of visual attention has the potential to improve the quality of the user experience. In particular, while not effective for all users, the user models proved effective in supporting the player interaction mostly improving the results achieved by the players.
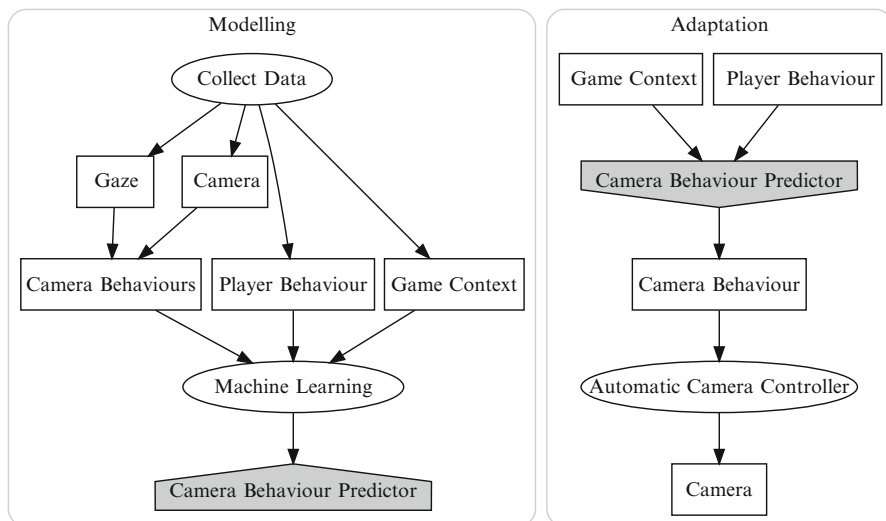
Fig. 11.4 Camera behaviour modelling and adaptation phases of the adaptive camera control methodology proposed by [17] (Image used with permission and adapted from [17])

Affective Cameras

Another fundamental aspect of player experience that has been studied in relationship to virtual cinematography is the player's affective state, both in terms of the ability of the viewer to understand character emotions and in terms of the effect of the cinematographic choices on the viewer's affective state.

Studies on emotions and cinematography in non-interactive media have investigated connections between low level cinematographic features and the viewer's experience; for instance, [57] studied the relationship between camera and object motion and the emotional responses of humans, concluding that an increase of motion intensity on the screen causes an increase in the viewer's arousal. Hanjalic and Xu [33] studied further the relationship between viewer's emotions and cinematography and developed a deterministic model of arousal and valence based on a combination of on screen motion, shots rhythm and sound energy. Sun and Yu [59] followed the same approach employing a non-deterministic method (i.e. Hidden Markov Models) to model the relationship between the aforementioned cinematographic features and four affective states: joy, anger, sadness and fear.

The first study exploring the role of emotions in interactive cinematography has been presented by [64], who prosed a bottom-up approach to shot definition based on a number of camera's affective states. In *CameraCreature*, there is no plan driving the movements of the camera, which is instead modelled as an agent moving in the virtual environment and reacting to the actions of the other agents. The agent controlling the viewpoint has an ethologically inspired structure based on sensors,

emotions, motivations and action-selection mechanisms. The camera agent shares this structure with all non-player characters (*Actors*) in the virtual environment and is able, through virtual sensors, to detect their emotions and the type of action they are performing. Tomlinson et al. [64] envision their work as *"a first step toward a future of interactive, emotional cinematography"*, which can be seen as an early definition of the concept of game cinematography. However, while the architecture proposed is potentially flexible enough to consider player emotions and actions, no directions are given on how this could be done.

Yannakakis et al. [67] studied the impact of camera viewpoints on player experience and developed a computational model to predict this impact, demonstrating the existence of a relationship between player emotions, physiological signals and camera parameters. However, the features employed in the model described cinematography using low-level camera parameters such as height or distance, which are unable to express the content of the visualised images. Burelli [12] performed a study that extended the aforementioned work by analysing the camera behaviour in terms of composition and by extending the analysis across different genres with richer game mechanics. The results confirm some of the findings revealed by [67], but there is evidence that the relationship between camera behaviour and player experience can be better explained by describing the cinematographic experience through more high-level features, such as shot spacing or symmetry, as these features allow us to understand what is the visual content that is reproduced on screen. Furthermore, the results reveal that the task the player performs affects the relationship between experience and visualisation.

Future Directions in Game Cinematography

Cinematographic games are a rising genre in the computer games industry and an increasing number of titles published include some aspects of cinematography in the gameplay or the storytelling. At the present state, camera handling in computer games is managed primarily through custom scripts and animations, and there is an inverse relationship between player freedom and cinematographic quality. However, the studies described in the previous sections show that there is a strong potential for improvement on the current state-of-art in game cinematography, especially towards building a better understanding of the impact of cinematography on player experience and how this could be leveraged to make better and new types of games.

Consequently, we see a number of future research directions that could be pursued. For instance, to foster a stronger awareness in the application of cinematography to games it would be extremely important to develop a taxonomy of game cinematography similar to the one that has been developed in classic cinematography throughout its history. Such taxonomy should be built following the approach delineated by [67] and [12] as this would allow game designers and developers to make choices on game cinematography with awareness of their impact on player experience and player interaction.

Furthermore, we envision a stronger interconnection of game cinematography with procedural content generation: this would allow to further develop the presentation aspect of game generation helping to move the field towards complete automatic game generation [63]. In order to do this, it would be necessary to extend the focus of game cinematography research towards a more general analysis of game presentation, including aspects such as visual aesthetics [44] and to study its perception and its effects on the user's understanding of the virtual world events. For instance, investigating how semiotics and narrative cognition can be integrated within the frameworks of automatic content generation and virtual cinematography, would allow the generation and the visualisation to become one coherent process which takes into account the signification of the content and the events [26].

Finally, following the trends in embodied artificial intelligence [20] and artificial intelligence for physical games [29], it could be possible to investigate the application of the results achieved in virtual game cinematography back to the physical world. For instance, thanks to the recent advancements in fields such as computer vision and robotics [35, 42, 48] and the introduction of ever more miniaturised filming equipment, micro unmanned aerial vehicle could be used as intelligent autonomous agents that could film a physical game for remote gaming or game broadcasting and recording.

Conclusion

In this chapter, we have presented an overview of the field of virtual cinematography and its application to computer games, and we defined what specifically characterised game cinematography. Moreover, we have analysed how the focus of the research in the field is shifting from pure algorithmic studies, focused on developing more robust and efficient algorithms to automatically animate the virtual camera, to studies which analyse different aspects of the relationship between visualisation, story and player interaction. In particular, in both game and traditional cinematography, there is a growing interest in understanding the impact on the viewer's cognitive processes and affective state of camera movements, editing and other cinematographic aspects. Finally, we have highlighted a number of possible future directions for game cinematography both towards a deeper understanding of the player's cinematographic experience and towards new applications of game cinematography beyond traditional computer games.

References

1. Arijon D (1991) Grammar of the film language. Silman-James Press, Los Angeles
2. Bares WH, Lester JC (1997) Cinematographic user models for automated realtime camera control in dynamic 3D environments. In: International conference on user modeling, Chia Laguna. Springer, pp 215–226

3. Bares WH, McDermott S, Boudreaux C, Thainimit S (2000) Virtual 3D camera composition from frame constraints. In: ACM multimedia, Marina del Rey. ACM, pp 177–186
4. Bares WH, Zettlemoyer LS, Rodriguez DW, Lester JC (1998) Task-sensitive cinematography interfaces for interactive 3D learning environments. In: International conference on intelligent user interfaces, San Francisco. ACM, pp 81–88
5. Beckhaus S, Ritter F, Strothotte T (2000) CubicalPath – dynamic potential fields for guided exploration in virtual environments. In: Pacific conference on computer graphics and applications, Hong Kong, pp 387–459
6. Bleszinski C (2007) Gears of War. Microsoft Game Studios
7. Blinn J (1988) Where am I? What am I looking at? IEEE Comput Graph Appl 8(4):76–81
8. Blizzard Entertainment (1998) Starcraft
9. Blizzard Entertainment (2004) World of warcraft
10. Bourne O, Sattar A, Goodwin S (2008) A constraint-based autonomous 3D camera system. J Constraints 13(1–2):180–205
11. Bungie Studios (2001) Halo: combat evolved
12. Burelli P (2013) Virtual cinematography in games: investigating the impact on player experience. In: International conference on the foundations of digital games, Chania. Society for the Advancement of the Science of Digital Games, pp 134–141
13. Burelli P, Di Gaspero L, Ermetici A, Ranon R (2008) Virtual camera composition with particle swarm optimization. In: Butz A, Fisher B, Krüger A, Olivier P, Christie M (eds) International symposium on smart graphics. Volume 5166 of lecture notes in computer science. Springer, Berlin/Heidelberg, pp 130–141
14. Burelli P, Jhala A (2009) Dynamic artificial potential fields for autonomous camera control. In: AAAI conference on artificial intelligence in interactive digitale entertainment conference, Palo Alto. AAAI
15. Burelli P, Yannakakis GN (2010) Global search for occlusion minimisation in virtual camera control. In: IEEE congress on evolutionary computation, Barcelona. IEEE, pp 1–8
16. Burelli P, Yannakakis GN (2011) Towards adaptive virtual camera control in computer games. In: Dickmann L, Volkmann G, Malaka R, Boll S, Krüger A, Olivier P (eds) International symposium on smart graphics, Bremen. Volume 6815 of lecture notes in computer science. Springer, Berlin/Heidelberg, pp 25–36
17. Burelli P, Yannakakis GN (2015) Adaptive virtual camera control trough player modelling. User Model User-Adapt Interact 25(2):155–183
18. Cage D (2010) Heavy rain
19. Charles F, Lugrin J-l, Cavazza M, Mead SJ (2002) Real-time camera control for interactive storytelling. In: International conference for intelligent games and simulations, London, pp 1–4
20. Chrisley R (2003) Embodied artificial intelligence. Artif Intell 149(1):131–150
21. Christianson D, Anderson S, He L-w, Salesin DH, Weld D, Cohen MF (1996) Declarative camera control for automatic cinematography. In: AAAI, Portland. AAAI Press, pp 148–155
22. Christie M, Normand J-M, Olivier P (2012) Occlusion-free camera control for multiple targets. In: ACM SIGGRAPH/Eurographics symposium on computer animation. Eurographics Association, pp 59–64
23. Christie M, Olivier P, Normand JM (2008) Camera control in computer graphics. In: Computer graphics forum, vol 27, pp 2197–2218
24. Core Design (1996) Tomb Raider
25. Dominguez M, Young RM, Roller S (2011) Design and evaluation of afterthought, a system that automatically creates highlight cinematics for 3D games. In: AAAI conference on artificial intelligence in interactive digitale entertainment
26. Eco U (1984) Semiotics and the philosophy of language. Indiana University Press, Bloomington
27. El-Nasr MS (2002) Story visualization techniques for interactive drama. In: AAAI spring symposium, pp 23–28
28. Enticknap L (2005) Moving image technology: from zoetrope to digital. Wallflower Press, London

29. Frazier SJ, Riedl MO (2014) Toward using games and artificial intelligence to proactively sense the real world. In: AI & GAME symposium, London. AISB
30. Haigh-Hutchinson M (2009) Real-time cameras. Elsevier
31. Halper N, Helbing R, Strothotte T (2001) A camera engine for computer games: managing the trade-off between constraint satisfaction and frame coherence. Comput Graph Forum 20(3):174–183
32. Hamaide J (2008) A versatile constraint-based camera system. In: AI game programming wisdom 4, pp 467–477
33. Hanjalic A, Xu LQ (2005) Affective video content representation and modeling. IEEE Trans Multimed 7(1):143–154
34. He L-w, Cohen MF, Salesin DH (1996) The virtual cinematographer: a paradigm for automatic real-time camera control and directing. In ACM SIGGRAPH, New Orleans. ACM, pp 217–224
35. He ZHZ, Iyer R, Chandler P (2006) Vision-based UAV flight control and obstacle avoidance. In: American control conference, Minneapolis. IEEE, pp 2166–2170
36. Id Software (1993) Doom
37. Irwin DE (2004) Fixation location and fixation duration as indices of cognitive processing (chapter 3). In: Henderson JM, Ferreira F (eds) The interface of language, vision, and action: eye movements and the visual world. Psychology Press, New York, pp 105–133
38. Jhala A, Young RM (2005) A discourse planning approach to cinematic camera control for narratives in virtual environments. In: AAAI, Pittsburgh, number July. AAAI, pp 307–312
39. Jhala A, Young RM (2010) Cinematic visual discourse: representation, generation, and evaluation. IEEE Trans Comput Intell AI Games 2(2):69–81
40. Kamiya H (2001) Devil may cry
41. Kojima H (1998) Metal gear solid
42. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, Lake Tahoe. Neural Information Processing Systems Foundation, pp 1097–1105
43. Lasseter J (1995) Toy story
44. Liapis A, Yannakakis GN, Togelius J (2012) Adapting models of visual aesthetics for personalized content creation. IEEE Trans Comput Intell AI Games 4(3):213–228
45. Lino C, Christie M, Lamarche F, Schofield G, Olivier P (2010) A real-time cinematography system for interactive 3D environments. In: ACM SIGGRAPH/Eurographics symposium on computer animation, pp 139–148
46. Martinez HP, Jhala A, Yannakakis GN (2009) Analyzing the impact of camera viewpoint on player psychophysiology. In: International conference on affective computing and intelligent interaction and workshops. IEEE, pp 1–6
47. McDermott S, Li J, Bares WH (2002) Storyboard frame editing for cinematic composition. In: International conference on intelligent user interfaces, San Francisco. ACM, pp 206–207
48. Meier L, Tanskanen P, Heng L, Lee GH, Fraundorfer F, Pollefeys M (2012) PIXHAWK: a micro aerial vehicle design for autonomous flight using onboard computer vision. Auton Robot 33:21–39
49. Mikami S (1996) Resident evil
50. Olivier P, Halper N, Pickering J, Luna P (1999) Visual composition as optimisation. In: Artificial intelligence and simulation of behaviour
51. Picardi A, Burelli P, Yannakakis GN (2011) Modelling virtual camera behaviour through player gaze. In: International conference on foundations of digital games, Bordeaux. ACM, pp 107–114
52. Pickering J (2002) Intelligent camera planning for computer graphics. PhD thesis, University of York
53. Ranon R, Urli T (2014) Improving the efficiency of viewpoint composition. IEEE Trans Vis Comput Graph 2626(c):1–1
54. Raynal F (1992) Alone in the dark
55. Remedy Entertainment (2001) Max Payne

56. Shaker N, Yannakakis GN, Togelius J (2010) Towards automatic personalized content generation for platform games. In: AAAI conference on artificial intelligence in interactive digitale entertainment
57. Simons RF, Detenber BH, Roedema TM, Reiss JE (1999) Emotion processing in three systems: the medium and the message. Psychophysiology 36(5):619–627
58. Soanes C, Stevenson A (2005) Oxford dictionary of English. Oxford University Press, Oxford/New York
59. Sun K, Yu J (2007) Video affective content representation and recognition using video affective tree and hidden Markov models. In: Paiva A, Prada R, Picard R (eds) Affective computing and intelligent interaction, Lisbon. Springer, pp 594–605
60. Sundstedt V, Stavrakis E, Wimmer M, Reinhard E (2008) A psychophysical study of fixation behavior in a computer game. In: Symposium on applied perception in graphics and visualization, Los Angeles. ACM, pp 43–50
61. Thawonmas R, Oda K, Shuda T (2010) Rule-based camerawork controller for automatic comic generation from game log. In: IFIP international conference on entertainment computing, Seoul, pp 326–333
62. Thorson E (1994) Using eyes on screen as a measure of attention to television. In: Lang A (ed) Measuring psychological responses to media messages. Routledge, New York
63. Togelius J, Champandard AJ, Lanzi PL, Mateas M, Paiva A, Preuss M, Stanley KO (2013) Procedural content generation: goals, challenges and actionable steps. In: Artificial and computational intelligence in games, vol 6, pp 61–75
64. Tomlinson B, Blumberg B, Nain D (2000) Expressive autonomous cinematography for interactive virtual environments. In: International conference on autonomous agents, Barcelona, p 317
65. Toyama K (1999) Silent Hill
66. Ware C, Osborne S (1990) Exploration and virtual camera control in virtual three dimensional environments. ACM SIGGRAPH 24(2):175–183
67. Yannakakis GN, Martínez HP, Jhala A (2010) Towards affective camera control in games. User Model User-Adapt Interact 20:313–340
68. Yannakakis GN, Togelius J (2011) Experience-driven procedural content generation. IEEE Trans Affect Comput 2:147–161

Chapter 12
From Sinewaves to Physiologically-Adaptive Soundscapes: The Evolving Relationship Between Sound and Emotion in Video Games

Tom A. Garner

Abstract This chapter examines the dynamic and evolving relationship between sound and emotion within the context of video games. How sound in games has been utilised to both infer and evoke emotion is discussed, commencing with an historical review that traces back to video games' humble beginnings. As we move towards the present day this chapter looks at how biofeedback technology, that can facilitate the control and procedural generation of game sound content by way of player-emotion, is transforming the lateral affective interplay between player and video game into something more circular.

The Power of Sound to Evoke Emotion During Video Gameplay

Sound is a critical component to consider when developing emotionality as it is directly associated with the user's experience of emotions [1, 56]. It has been posited previously that sound carries more emotional content than any other part of a computer game [50]. Grimshaw and colleagues [29] discovered that players felt significant decreases in immersion and gameplay comfort when audio was removed from gameplay; an assertion also made by Jørgensen [37] who, via observations and conversations with players, revealed that an absence of sound caused a reduction in engagement such that "the fictional world seems to disappear and [. . .] the game is reduced to rules and game mechanics" (3). There are many acoustic and psychoacoustic properties of sound that could be investigated as to their affect-inducing potential. Some are quantitative, in that they can be objectively measured and applied to synthesis and audio processing. For example, Cho and colleagues [13] provided evidence that pressure level, loudness and sharpness of a sound can directly affect emotional valence and intensity. Other properties are more qualitative

T.A. Garner (✉)

School of Creative and Cultural Industries, University of Portsmouth, Portsmouth, UK
e-mail: tom.garner@port.ac.uk

and relate to the interpreted meaning of sound. They can be influenced by factors such as culture, experience, context and expectation.

Based upon the recent developments in games technology, there is certainly an argument to be made as to the significant power of sound to facilitate affective gaming by way of immersion. Sound possesses an innate ability to surround the player as it infuses into (and resonates around) their physical space. The resultant immersion fundamentally supports the affective potential of game content as the process of bringing the player and the game world closer together imbues the content with much greater meaning [36, 64]. The entertainment technology industries have made several attempts to equal this particular strength of sound in visual hardware. Recent attempts at vision-based immersion (that include curved screen technology, revivals of 3D television and virtual reality headsets such as the forthcoming *Rift* [49]) show that there is a great deal of investment going into generating an immersive visual experience, to try and match that which has already been achieved by sound.

8-Bit Affect: Sound and Emotion in Video Games' Formative Years

Audio Technology and Its Sound Design and Composition Affordances

In 1958, William A. Higinbotham embarked on one of the very first documented attempts to configure a games application within a computer system. The Systron-Donner analogue computer, primarily intended for executing military aerospace applications, was deemed by Higinbotham to be lacking in dynamism and limited in its ability to garner attention from research investors. To that end, he created *Tennis for Two*, a rudimentary tennis simulation displaying full-motion graphics by way of an oscilloscope (see [59]). The game was silent and subsequent games systems would continue to be as such until 1971. The next landmark event occurred in 1961 with the development of what is typically referred to as the first official video game, *Spacewar*! The game ran upon the MIT developed PDP-1 (Programmed Data Processor-1) and was intended to demonstrate the full range of the system's processing capabilities when pushed to the limit [27]. Whilst the PDP-1 did indeed have the capacity to generate audio signals and was utilised for this purpose a couple of years later (see [10]), Spacewar! did not incorporate sound. The silence continued as games systems invaded the home and the Magnavox Odyssey became the world's first commercial video games console. As the PDP-1 before it, the Odyssey did not incorporate audio into its games [58].

In 1971, arcade games went sonic with the advent of *Computer Space* (Nutting Associates) followed a year later by Atari's *Pong* (1972). Whilst the former of these titles remains the foremost sound enabled game, its relatively weak popularity

with audiences has arguably made Pong the more significant entry in the history of game sound – with some sources even citing (albeit incorrectly) it as the first ever game to produce sound. Pong presented three discrete monaural tones (sine waves of varying frequency and duration) to coincide with in-game activity. The limitations of the prototype system on which Pong was being developed meant that no single component or section of the circuitry was intended to generate sound. This required the engineers to creatively wire the system's sync pulse generator to an active speaker, enabling a component originally designed to determine aspects of the visual output to simultaneously generate an accompanying sound [65]. The subsequent step change occurred in the early 8-bit era of gaming, in systems such as the Bally Astrocade (1977). This system demonstrated the first game to exploit a microprocessor as a means of generating sound for a game, *Gunfight* (Taito 1977). The game utilised a single 3-voice channel plus a noise generator and a vibrato modulator [63] which enabled a white-noise burst to sound each time the player's avatar fired their pistol and for two short monophonic melodies to sound individually – one when the player successfully shoots the adversary, and another when their avatar is killed.

The years that followed were undoubtedly dominated by Atari who in 1977 released their Video Computer System (commonly known as the 2600) before unveiling the seminal *Space Invaders* (Atari 1978). The 2600 utilised the then advanced Television Interface Adaptor (TIA) chip that provided two individual monaural channels, each able to produce 4-bit (16 values) volume control, up to 32 different pitches and 16 alternative sound registers. This allowed the 2600 to run games that performed both music and sound effects simultaneously and also produced audio of a greater dynamic and compositional range. The 2600 was also one of the first systems to enable the generation of square waves, giving the 2600 a very distinctive tone and founding the long standing affection many have for the 'chip tune' sound [12]. A further development of significance arrived courtesy of Magnavox's sequel gaming system, the Odyssey^2 (1978). The Odyssey^2's integrated audio hardware was limited (providing only a single monaural channel of audio) in comparison to Atari's 2600. However, an expansion module, *The Voice*, could be plugged into the console cartridge port to facilitate voice synthesis [5]. The Voice module's hardware included an integrated speaker and volume control, playing only games built specifically to utilise the module. This enabled the system to not only produce a much wider array of timbres, giving each game a more distinct sonic quality, but also by way of the speech synthesis allowed the game to communicate audible speech to the player [41].

As 8-bit moved into the 1980s, Atari's 2600 successor, the 5200 (1982), demonstrated that games audio technology was continuing to improve. The 5200 utilised the POKEY (Pot Keyboard Integrated Circuit – admittedly a rather loose acronym) audio chip, facilitating four channels of audio and enabling multiple strands of sound to be performed in parallel. The volume, frequency and waveform of each channel could be modified independently and the chip also incorporated a high pass filter. Consequently, music and sound effects could become more diverse

and richly detailed as POKEY enabled full musical polyphony in tandem with sound effects and the ability to compose music based upon the 12-tone equal temperament scale.[1] This meant that game music could include a much greater range of melodic and harmonic constructions (such as transposition) and accurately recreate classical and contemporary/popular music pieces [4].

1983 saw a noteworthy development for game sound, this time not drawn from advances in microprocessor technology, but from the media format. *Dragon's Lair* (Cinematronics 1983) originally debuted on arcade machines and played from a Laserdisc as opposed to traditional cartridges. The hugely superior data capacity of the Laserdisc allowed for pre-rendered sound effects and musical content to be auditioned during gameplay. This meant that the aesthetic freedom of the sound designers/composers was widened to the extent that almost any conceivable soundscape could be created, from a symphonic orchestral swell to sound effects that rivalled motion pictures. As the precursor to the Compact disc, the Laserdisc was arguably, after a slight divergence in the 16-bit era, a sign of things to come.

Gaming largely returned to cartridge media in the 16-bit era and focussed upon improving the quality of microprocessor-based sound synthesis. However, this console generation is where we also observe the proliferation of sampling technology that enabled systems to combine synthesised and pre-recorded audio samples as a means to improving overall quality. This fourth generation of consoles was largely dominated (to varying degrees of success) by the TurboGrafx-16 (NEC 1987), the Neo Geo AES (SNK 1991), the Mega Drive/Genesis (Sega 1988), and the Super Famicom/Super Nintendo Entertainment System (Nintendo 1990). All of these systems utilised comparable audio systems that incorporated much improved synthesis capability over the prior 8-bit machines. This included a greater number of audio channels, multiple channel types and introduced frequency modulation synthesis, a technique that could generate more complex waveforms and therefore produce a wider range of timbres). With regards to sound design, FM synthesis greatly increased the number of musical instruments that could be synthetically represented to an acceptable quality level. Keyboards (piano and organ), pitched percussion and plucked instruments (guitar, pizzicato strings, etc.) were particularly well received and as a result featured prominently in many 16-bit games (see [15]: 38). Additional 16-bit sound features included pitch-bend support and triangle-wave channels that provided further expansion of the sound designer's toolkit, enabling an even wider range of timbres and modulations to be performed [12].

All of the above 16-bit systems supplemented their synthesis capabilities with alternative forms of pulse-code modulation (PCM), bringing digital audio to home consoles and enabling sampling to become part of the sound designer's arsenal. The primary differentiation between these systems can be observed in the number of PCM channels and in each channel's sampling rate as the developers had to

---

[1]A form of tuning that spaces out each note within the octave equally along the frequency spectrum.

balance high channel number (rich, detailed soundscapes) against high sampling rates (high quality individual sound samples). In terms of sound control, 16-bit systems were able to facilitate increasingly complex sound processing effects. For example, the Super Famicom supported full envelope control of waveform dynamics (designers could precisely control the attack, decay, sustain and release [ADSR] of a sound), Gaussian interpolation (a processing technique that enabled the upscaling of sampled audio to a higher sampling rate as a means of improving quality without greatly increasing processor demands) and convincing echo effects. Within this generation we also see the highly significant development of surround sound beginning to appear, with Dolby Pro-Logic enabling the power to spatialise game sound to an unprecedented degree [57].

As we progress into the fifth generation of 32-bit home console gaming there is a distinct favour shown towards the compact disc as game media (acknowledging of course Nintendo disrupting the order on two counts, both by entering a 64-bit system into the mix and by continuing to employ the cartridge). Of the consoles that populate this generation, the PlayStation (Sony 1994, commonly known as the PSX) arguably rose to become the dominant force. Like the preceding generation of consoles, the PSX's audio hardware included a standalone sound processing unit but with a doubling of processor memory from 8-bit to 16-bit plus a 44.1 kHz sampling rate – effectively equalling the quality of commercial music CDs. The impact of switching primarily to pre-rendered audio significantly changed the sound designer's working process as they were no longer required to have a working knowledge of games architecture and sound programming. Instead they could produce the desired music and sounds independently of the games system, then the samples would be programmed in, typically by a separate developer [12].

The PSX sound processing unit supported advanced features that elevated it notably from prior systems. In addition to full ADSR control, the PSX supported real-time digital looping and reverb to enable more flexible control over pre-recorded samples. Whilst these improvements to digital audio control were significant, the PSX sound processor was not solely reliant on samples and also incorporated synthesis (specifically MIDI performance). This continued the responsibility of game sound designers to balance the pros and cons of each output type. However, unlike in the previous generation, which required a compromise between quality and processor resources, the PSX trade-off was primarily between quality and interactivity. Whilst pre-rendered digital audio enabled convincing performances of elements such as musical instruments and speech, it provided very little real-time flexibility. By comparison MIDI could be much more responsive to the interactive nature of a game due to the fact that MIDI information consumed much less disc space than samples, enabling many times more audio content to be stored within the game media. Naturally, more available content could reflect a greater number of game states, objects, events and transitions, thereby supporting a much more responsive game sound environment. As a result, the way in which the PSX's audio hardware is implemented with regards to music, speech and sound effects is significantly different from one game to another.

Evoking Player Emotion Via Sound in the 8, 16 and 32-Bit Eras

The continuing evolution of game sound technology brought on an increasingly wide array of compositional and sound design techniques as the limitations that had plagued early generation designers were gradually lifted. These developments to the craft of sound and music design present a most interesting frame from which we can examine the changing approaches to evoking emotion by way of sound and relate these approaches to auditory processing and emotion research.

The creation of *Pong* (1972) saw some of the first sounds in a video game, with basic sine waves representing the ball hitting the paddles, ricocheting off of the screen edges and passing the paddle as one player scored a point. Each action was distinguished sonically by giving each event a sine wave of alternate frequency. This basic sound design had little capacity to evoke player emotion by design but the intention for such was certainly present, with Atari originally requesting that when the player lost a point, a booing sound would play, with the intention of stimulating frustration and competitiveness in the player in order to encourage their continued engagement [65]. Examining several early-generation games with regards to evoking emotion by way of music, it is apparent that established concepts of musical affect are characteristically employed to evoke the desired emotional response as dictated by the gameplay. For example, returning to *Gunfight* (1977), when the player is shot and killed, the Funeral March by Chopin can be heard, evoking the sensation of tragedy by way of exploiting semantic association (via the cliché of the theme) and utilising a melody constructed from a minor key (see [34]: 2).

Moncrieff and colleagues [45] reference attack-decay-sustain-release (ADSR) as a quantifiable acoustic parameter that presents a significant correlation between sound and specific emotional responses. This is in the context of broader dynamic changes across an overall sonic environment. Specifically, the research suggested that a period of silence (or very low overall intensity) followed by a sharp attack and quickly releasing back to silence was characteristic of a startle response, whilst a slow attack from low to high intensity that sustained at high was reliable for creating suspense. This is further supported by Xu and colleagues [67], who suggest that an audio shock is most effective when it is preceded by silence. These two techniques can be observed in various games across the generations, from the Scissor-man jumping from the bathtub in *Clock Tower* (Human Entertainment 1995) to the slowly increasing loudness and harmonic tension of the strings as the player delves into the dark recesses of *Silent Hill* (Konami 1999).

The release of *Space Invaders* (1978) is an excellent example of manipulating periodicity (tempo) to elicit substantial affective change in players. As the alien hordes descend towards you and the likelihood of 'Game Over' increases, so does the tempo of the background music. This closely relates to the concept of entrainment (see [1]) in which, by way of anxiety, the increasing musical tempo causes the player's heart rate to increase. Parker and Heerema [50] suggest that an evolutionary survival instinct exists today that encourages humans to associate

low-pitched sounds such as growls and rumbles with predators and, in response to such sounds, are likely to experience fear. They also suggest that, by way of a social association, human sounds that imply a horrific act (such as a scream) evoke fear by way of an inherent empathic response. Exploitations of these concepts can also be seen in earlier generations of gaming, such as the deep roar of the titular antagonist character in *Sinistar* (Williams Electronics 1982) or the scream of Lara Croft as she falls from too great a height in *Tomb Raider* (Eidos Interactive 1996). With regards to early 8 and 16-bit games, implementing the pentatonic scale for an earcon (a short musical phrase that forms part of the auditory display by consistent association with a particular game event – see [33]) to generate a positive affect is a common practice and can be exemplified by the 'level complete' sound of *Super Mario Bros* (Nintendo 1983) and in *Sonic the Hedgehog* (Sega 1991), where the three-tone 'sparkle' sound plays when a ring is collected. These examples do not present an exhaustive list of techniques that reflect research into sound and emotion but they do indicate that video games have consistently been a medium for affective sound practices throughout their history. The following section moves us towards the present day and reflects upon the more cutting edge approaches to eliciting player-emotion by way of sound.

The Contemporary Sound-Emotion Relationship

With the technical resources currently available within video games systems, the separation of audio programmer from sound designer/composer has reached its logical conclusion with musicians, composers and foley artists[2] now being brought in from outside the video games industry to create content [12]. Sound now enjoys research attention as a key component of the current developments being made in procedural content generation, with sonification algorithms being developed that can control various granular aspects of sound and music (instrumentation, pitch, intensity, etc.) in real-time [42].

Modern Technical Developments in Games Audio Technology

As home consoles 'levelled-up' to the eighth generation with the release of the PlayStation 4 (Sony 2013), Xbox One (Microsoft 2013) and Wii-U (Nintendo 2012) audio technology has been significantly enhanced. Several noteworthy developments have been accomplished that enable sound designers to create affective content that bridges theory and technical possibility more than ever before. For example, the PlayStation 4 utilises the *TrueAudio* [3] audio processing unit that

---

[2]A designer responsible for recreating natural sounds in a studio.

enables advanced digital signal processing (DSP) effects that include convolution reverberation (a virtual recreation of an actual acoustic space by way of a stored impulse response sample) – a processing effect that enables real-world acoustic phenomena (e.g. the dampening/reflecting properties of different environmental materials) to be more accurately modelled. With this effect, the player could walk down a virtual corridor engaging in conversation with an AI and as they move from a carpeted room to a polished wooden floor the audio engine can create a distinct transformation in the timbre of the voices to match the environment. The increased processing power available facilitates real-time control of multiple DSP effects across numerous audio channels simultaneously, meaning that the virtual acoustic environment can respond to game metrics and player actions to a much greater degree than on earlier systems. Refinements made to multi-channel spatialisation algorithms enable game sound events and objects to be easily discernible across all three spatial axes, enabling players to more accurately perceive the position of sounds be they above, below, in front, behind, proximate or distant. These real-time processing capabilities also extend to music and modern games typically now incorporate adaptive scores that emotionally match the game environment/situation with a particular theme, variation or stab.[3]

Contemporary Approaches to Affective Experience in Gaming via Sound

With the seemingly almost limitless potential of modern gaming technology to craft soundscapes to a highly precise requirement, game sound design now has the capability to reflect contemporary research perspectives with techniques that are more complex, responsive and adaptive. For example, in a recent article Karen Collins [16] explores how to make gamers cry, primarily suggesting that the solution lies in creating an embodied interaction with music and sound as a means of generating emotional involvement. Collins refers to the concept of mirror neurons: the notion that the same neural pattern will potentially fire when an individual observes (or in our case, hears) an action as will fire when they perform it. Collins expands upon this concept to propose that game sound that is sufficiently tied, at a semantic level, to a physical action will evoke an emotional response comparable to that which would arise should the individual actually be executing that action. A good example of this can be found in *Tomb Raider* (Eidos Interactive 1996) when Lara Croft emerges for air after a prolonged period underwater and we hear her gasp for air. The mirror neuron theory would suggest that part of the reason we experience an intense sense of emotional relief upon hearing this sound is because we associate it with a sound we ourselves have made in a comparable scenario (as

---

[3]A singular, short and intense note or chord typically utilised to add dramatic impact.

anyone who has competed with childhood friends in a lung capacity competition at their local swimming pool will understand).

Research has also indicated that sound and emotion are connected at a behavioural level. Going all the way back to the original Pavlovian experiment it could certainly be suggested that the conditioned response to the ringing bell had a significant emotional component. More recent research has examined this within a gameplay context, revealing that sound has great potential to modulate emotional arousal and stress levels by way of positive reinforcement. An experiment by Dixon and colleagues [20] observed the behaviours of participants playing a digital video slot machine. They revealed that when a particular sound was conditionally associated with a win, performing that same sound in response to a loss often resulted in the participant experiencing a positive emotional sensation despite losing and, at the end of play, significantly overestimating the number of wins they had experienced during their session. An almost reverse of this effect can be observed in *Shadow of the Colossus* (Team Ico 2005) in which, after successfully defeating Valus, the first colossus in the game, a prolonged silence followed by a haunting refrain is performed as the creature falls. As a result, the player is initially positioned to feel a sense of achievement and pride as they have essentially won the game, but the significant shift in tone (as established by the score) creates a jarring change in emotional state as the player considers that their victory is in fact a rather profound loss.

Another concept suggests that the association of sounds or musical themes with particular game elements, such as characters or locales, can bestow particular game sounds with significant affective potential [60]. *Left 4 Dead* (valve 2008) exploits this concept as a core aspect of the gameplay by endowing each of its 'special' adversaries with both a highly distinctive vocal timbre (such as the bloated bowel lurch of the Boomer or the chocked, scratchy scream of the Smoker) and a leitmotif[4] of comparable timbre to the vocal. These sounds all play when the particular character is close by but not yet visible on screen, creating an acousmatic effect that, upon repeat audition, creates an immediate emotional response in the player. This then causes a direct change in their behaviour as they adjust their combat strategy to manage the particular threat being indicated. In a review by Cole [14], two games renowned for their affective content, *Ico* (Team Ico 2001) and *Shadow of the Colossus* (Team Ico 2005) are analysed with regards to their sound content. Cole notes that the characters within these games all display unique sonic profiles as a means of establishing their personality and prompting emotional engagement from the player. The review also comments upon the power of sound in absentia, positing that well-placed silence can have an emphatically emotional impact. With regards to Ico and Shadow of the Colossus, Cole argues that the use of silence and minimalist soundscapes contributes particularly to a feeling of loneliness and isolation.

With the various parallels that exist between video games and cinema it is to be expected that certain auditory concepts that originated from film studies have been

---

[4]A short recurring musical phrase that has a designated association with a locale, event or character.

applied to video games. For example, Connor [17] discusses non-linear sound in film as an effective approach to evoking fear by way of manufacturing perceptions of animals in distress, citing the coupling of Janet Leigh's scream with Bernard Herman's string motif for the renowned shower scene from Hitchcock's *Psycho* (1960) as a prime example. The term 'non-linear' applies to sound when excessive amplitude (typically greater than that which could be produced by the vocal chords of an animal) causes the waveform to self-distort (by way of greater waveform peaks than troughs) and gives the sound a raspy timbre (see [31]). This technique is arguably prominent in video games such as *Amnesia*: *The dark descent* (Frictional Games 2010) in which various sounds are heavily amplified until they distort, then digitally compressed to avoid excessive decibel levels whilst retaining the non-linear timbre. Another affective sound technique that has a heritage in cinema is hyperreal sound. This term accommodates a range of techniques that range from the digital post production of sounds (expanding the dynamic range, compression to increase the 'punch', adding reverberation or delay effects, etc.), to the blending of multiple source sounds to create a new 'hybrid' sound. A great example of the latter is the colossal roar of the tyrannosaurus rex in *Jurassic Park* (1993) which began as the cry of an infant elephant, and the barking between velociraptors for which copulating tortoises provided the foundation ([8, 22]: 287). With regards to video games, it has been asserted that extended exposure to fictitious 'Hollywood' sounds has determined that genuine source recordings (shotgun blasts, footsteps in the snow, etc.) in a game would be perceived as 'flat' and 'lifeless', with limited capacity to immerse the player or evoke an emotional experience [56], suggesting that the majority of modern game sounds are hyperreal by default.

Various facets of video games facilitate emotion types that are common to film (see [61]), such as the *F-emotion* (fiction emotion: empathetic states also referred to as 'witness emotions' because they arise from the individual's observation of the fictional environment/scenario) and the *A-emotion* (artefact emotion: derived from appreciation of the artistry/craft that built the observed fiction, felt during brief realisations that the film/game is not real). The interactive nature of video games generates further emotion types in the *R-emotion* (representative emotion: denoting states evoked from action/interaction within a fictional world, see [61]) and the *G-emotion* (gameplay emotion: emotions that arise from a hybridisation of the F-emotion and R-emotion types, see [51]). For example, when undertaking the role of Gordon Freeman in *Half-life 2* (Valve 2004), we may witness the dystopia within which we are placed as we read the headline of a newspaper cutting pinned upon a board "Earth surrenders" (F-emotion). Stepping briefly outside the fictional world, we may then admire the game's graphical quality and the great level of detail that the cutting reflects (A-emotion). We may also simultaneously reflect on the fact that this environment is a direct consequence of our own prior actions (R-emotion) and that in our continuing play of the game it is our responsibility to produce further action to set things right (G-emotion).

Fear is a discrete emotion that has received particular research attention with regards to sound and video games. Many such articles specifically examine the spatial aspects of sound and their potential to evoke a fear sensation during

gameplay. Localisation, specifically techniques that obscure the position of a sound object (soundwave source) has been posited as critical when building suspense and terror [7]. Building a lo-fi audio soundscape consisting of many interfering sounds can increase disorientation and make cognitive processing significantly more difficult which can decrease the player's ability to cope with the in-game threat, consequently priming the player to experience fear [23]. Utilising reverberation and delay effects to increase positional ambiguity, particularly upon acousmatic sounds is a very well established technique in contemporary horror games (see [50]).

The techniques documented above all focus upon utilising the game engine exclusively as a means of controlling the soundscape and, consequently, player emotion. In the instances, much of the content is designed to be fixed within the game, creating what is often referred to as a cinematic experience. Whilst such approaches are popular within contemporary gaming there is arguably a significant interest in crafting more interactive, bespoke and responsive games content by way of more procedural practices that can respond to more than just the player's position and behaviour within the game. The next section discusses biometric interfaces as a strong candidate for supporting procedural sound generation in games and also looks at biofeedback techniques that have already begun to realise their potential for interpreting player emotion as a means of altering the game environment – essentially creating a loop in which the game influences emotion and, equally, emotion influences the game.

Psychophysiology and Biometric Game Control Interfaces

One significant new development with regards to contemporary games technology is the implementation of biometric equipment intended to measure physiological state in players and relay that information to the game engine as a means of influencing game parameters. Our interest remains on sound content within video games but we begin this section with a brief review of general biometric game applications before examining sound specifically. Psychophysiology refers to study of the relationships that exist between physiological events and psychological (cognitive, emotional and behavioural) processes [9]. Biometrics typically denotes authentication processes, utilising physiological characteristics within security and identification contexts (see Woodward et al. [66]). However, within the context of video games, biometrics more commonly associates more closely with the definition of psychophysiology and is utilised primarily as a means of gauging user experience information by way of an emotion-interpretation framework that attempts to predict affective and cognitive state from physiological data (see [44, 47]). Whilst biometric technology has yet to be incorporated directly into home consoles, the advent of motion-detection hardware such as the *Xbox Kinect* (Microsoft 2010) and *PlayStation Move* (Sony 2010) is indicative of the commercial and consumer interest in innovative video games technologies. Motion-detection technology has generated associations with video game sound in both research (for example Yoo et al. [69], who explore the use

of the Kinect to enable gesture-controlled sound creation) and commercial games such as *Child of Eden* (Ubisoft 2011) in which the player uses gesture to manipulate the game's sound as a central element of gameplay.

User experience testing is by no means the limit for implementations of this technology and much interest has begun to grow with regards to potential biofeedback applications in games. Recent research reveals this interest by way of exploring various benefits of game biofeedback systems, including: controlling attention and relaxation in people suffering from Attention Deficit Hyperactivity Disorder [2], modulating game control mechanics to facilitate deeper immersion [39] and altering in-game content as a dramatic device [48]. This section will first outline the specifications of the most commonly employed biometrics in gaming before examining how this technology is being implemented (both with regards to research and commercial video games development) and what these advances mean for the sound-emotion relationship in games.

Principles and Mechanics

One prominent principle of psychophysiology states that, when attempting to accurately predict psychological states from physiological signals, multiple types of measurement should be acquired in parallel as this will enable one measurement to verify the other [21]. As such, much biofeedback research implements multiple measures to the extent that we cannot review them all within this space. Therefore we shall focus upon the biometrics that most consistently feature within contemporary research as the primary measure, namely electrodermal activity (EDA, commonly known as skin response) and electroencephalography (EEG).

Electrodermal activity makes associations with psychological state by way of the peripheral nervous system that connects the extremities of the human body to the spinal cord and brain (see [47]). Looking at functional rather than regional aspects of the nervous system, EDA has been attributed to the sympathetic nervous system (a set of autonomic processes that govern excitation and prepare the body for fight-or-flight action) and is consequently associated with arousal [53]. EDA sensors detect the conductance levels of the skin as they alternate as a result of variations in sweat production from the eccrine glands [40], with increased sweat lowering the resistance levels of the skin thereby increasing electrical conductivity [55]. Essentially, emotionally arousing stimuli activates the sympathetic processes of the nervous system, which in turn activates the glands; releasing greater volumes of sweat and leading to increased electrical activity which is then measured by the biometric hardware (see [25]). The association between EDA and arousal is well established and EDA is currently the most commonly employed measure being utilised as a means of quantifying arousal [54]. Specifically as a measure of emotion-related arousal, limbic brain regions have been associated with EDA [68] and more recent research has argued that both affective and cognitive processes can be associated with EDA [18].

Electroencephalography (EEG, often referred to as a brainwave measurement) also examines electrical activity, but this time directly from the brain itself. Electrodes are placed in particular arrangements across the scalp and detect the voltage oscillations that result from neural activity. The difficulty of intentional user-manipulation or suppression of brainwave activity positions EEG as an excellent biometric for examining subconscious processes and subverting attempts by the user to present a false response [46]. EEG is often compared against brain imaging techniques such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) and is evaluated to be superior for video games studies due to its much greater portability, ease of use, temporal resolution (able to record thousands of states per second) and affordability [30]. Much like EDA (and much unlike fMRI and PET), the process of measuring EEG is relatively un-distracting as the equipment produces very little noise, does not obscure vision and is comfortable enough for the wearer to retain focus upon the test stimuli [24]. Whilst EEG is not without methodological difficulties, developments of this technology are continuing to reduce or overcome such issues. For example, the spatial resolution of EEG is relatively low in comparison to other imaging techniques, limiting our confidence when attempting to correlate specific regions of the brain and particular psychological processes. However, reducing the size of EEG sensors (increasing the number of sensors that can be fitted into an EEG cap) and refining the algorithms that process the EEG signal significantly improves spatial resolution [62]. An increasing volume of research experiments have drawn associations between emotion and EEG, with it being asserted that EEG has the potential to not just detect the presence of an emotion but to identify the discrete class of emotion such as happiness or fear (see [35]). Classification of human emotion within a video game context has also been examined with connections made between specific patterns of EEG activity and player-frustration [26] and also perceived levels of difficulty [11].

Utilising Biofeedback to Connect Player-Emotion to Game Sound

Biometric development reaches beyond our understanding of the affective potential of game sound to the concept of emotion-biofeedback loops: automated systems capable of amalgamating physiological responses with game state data (from individual events to overarching situations and surrounding virtual environments) to accurately and reliably infer emotional states during computer gameplay and feed that information back into the system. The game engine can then respond with changes to, potentially, any conceivable parameter of the game; from generating a sunrise in response to a player's happy state, to increasing the avatar's physical action statistics (run faster, jump higher) in response to a player's aggressive state. If such a system were utilised within modern games titles: increased concentration could enable the bullet-time function in *F.E.A.R* (Monolith 2006) or *Max Payne*

(Remedy 2001), elevated relaxation could increase your speed and chance of success in defusing a bomb in *Rainbow Six*: *Vegas* (Ubisoft 2006), and an angry emotional state could unlock additional 'renegade' conversation options in *Mass Effect* (Bioware 2007). Game sound within biofeedback systems has received comparatively little specific attention but it often remains a component of the game content that is manipulated by the affective interpretations received from biometric data (see [19]). Headlee and colleagues [32] do present a sound-centric, multiplayer biofeedback game in which electrocardiogram (ECG) data is used to control the overall soundscape quality, as players attempted to regulate their heartbeat to transition from an intense urban soundscape to one that is more relaxed. Their findings indicated much potential, as players were able to successfully utilise the biometric controls to achieve game tasks. Headlee and colleagues do concede that control using biometrics is difficult and inconsistent but also state that there is much opportunity for future improvement.

Whilst EEG has found favour as a means of discrete emotional classification, EDA has become well established as a reliable indicator of affective intensity. Changes in musical expression have been shown to reliably produce spikes in EDA [38] whilst non-musical sounds (sound effects) that were identified qualitatively by listeners as either emotionally intense or flat, reliably produced correspondingly high or low skin conductance levels [6]. In an experiment by Dekker and Champion [19] heart rate and EDA hardware was integrated into the *Half-life 2* (Valve 2004) game engine, enabling physiological indicators of emotional arousal to control particular aspects of the game's content. For example, when the player was in a low arousal (bored/calm) state, this was detected by way of a drop in heart rate and skin conductance which, when these measures reached a pre-set threshold, communicated with the game engine to perform a virtual proprioceptive heartbeat loop (and spawn an additional enemy for good measure). Recent research has also examined cardiovascular activity (specifically heart rate and heart rate variability) as a means of computationally modelling typical gameplay affective states such as fun, challenge, boredom and excitement [52]. This is, of course, not an exhaustive list of biometrics/biofeedback in games research. However, it does highlight common themes and preferences with regards to the biometric measures utilised and the particular emotions investigated. At the time of writing, such technology is yet to be implemented into the home console market and remains primarily a research and development area rather than an established commercial technology. Despite this, there is much reason to believe that the near future will bring many significant changes that will transform the nature of both affective gaming and game sound, as we briefly discuss in the closing section of this chapter.

The Future of Sound and Emotion in Video Games

Biometrics within video games is currently at a developmental stage but it is anticipated that within 10 years the technology will become mainstream; the central application of this technology being an adaptive system, capable of learning the

preferences, motivations, and emotional temperaments of individual players and with that information creating unique and evolving gameplay experiences [43]. Biofeedback systems facilitate the potential for radically improved artificial intelligence that could empower non-player characters (NPC) with emotional intelligence. NPCs could react appropriately in real-time to players' emotion states, enabling the player to interact with characters like never before and simultaneously opening up a world of possibilities for new game mechanics. Biofeedback has the potential to allow a player to: intimidate or calm a suspect during an interrogation, barter with a passing traveller over the cost of a new plasma rifle, or convince a friendly character to believe in and join your crusade, all by way of feeding physiological information into the game engine that is then translated during gameplay into an appropriate NPC response. This has particular exciting potential for all aspects of game sound including how speech, sound effects and music are presented. Empowering NPCs with empathic abilities by way of biometrics could enable them to generate conversational responses perfectly nuanced to the player's affective state. As the game detects the players attention levels by way of EEG and pinpoints the subject of their attention via eye-tracking, the mix and equalisation of the soundscape could adapt in real-time to emphasise the sound object under focus, bringing that particular object into the auditory foreground. Adaptive musical scores could respond to the player's biometric output by mirroring certain timbral and rhythmic qualities of their physiology (such as their heart beat or respiration), or based upon the system's affective interpretation, regulate the emotional content of the music to either support a kind of emotional homeostasis (an attractive proposition for those who enjoy survival horror games but find them too intense) or to try and push the player to extreme intensities of emotional experience.

At present, we are approaching an exceptionally exciting point in affective video game sound as the progress of technology has enabled the realisation of highly complex concepts and frameworks of acoustics and auditory processing. Furthermore, such concepts are themselves continuing to develop as we refine our understanding of sound as a phenomenon. We are beginning to consider sound within contemporary psychological perspectives as an embodied experience, determined not only by the material acoustic component of sound (the soundwave), but also by countless factors within the mind, the body and the environment [28]. For video game sound this type of perspective is highly integrative and promises to support richer gameplay experiences in which sound content is deeply connected to both the player and the game. By way of biofeedback technology, this perspective could be realised in future games to create a deeply immersive and emotionally engaging experience. As you (the player) walk across the room and approach the cellar door, you fear for what you may find on the other side. As you move in closer a faint scratching from behind the door makes your palms begin to sweat. The EDA sensor feeds this information into the game, which contextually detects the emotion by cross-analysing your physiological change of state with game data. A distorted string motif begins to crescendo and the rhythm of your heart beat is matched perfectly by the score's underlying percussion. The door creaks open and you stare into the darkness, scanning the environment. A shrouded figure just visible in the corner of the room stands motionless and does not move until the event related

potential, generated the moment that your attention is directed to the figure, causes the figure to lurch forward at the perfectly terrifying moment. As you try to defend yourself, the figure removes their shroud and identifies as a friendly AI. You realise it is all a prank and begin to relax; the game detects this and so the AI 'knows' this. They then stop trying to calm you down but all of a sudden, begin to look concerned, noticing that you've suddenly become terrified again. They ask you why you're still afraid; it was only a joke after all. Unfortunately you've spotted something else lurking in the darkness, moving closer . . .

References

1. Alves V, Roque L (2009) A proposal of soundscape design guidelines for user experience enrichment. Audio Mostly 2009. September 2nd-3rd, Glasgow
2. Amon K, Campbell A (2008) Can children with ADHD learn relaxation and breathing techniques through biofeedback video games? Aust J Educ Dev Psychol 8:72–84
3. AMD (2015) TrueAudio. http://www.amd.com/en-us/innovations/software-technologies/trueaudio. Retrieved 08 July 2015
4. Atari POKEY data Sheet (2015) http://krap.pl/mirrorz/atari/homepage.ntlworld.com/kryten_droidAtari/800XL/atari_hw/pokey.htm. Retrieved 01 July 2015
5. Boris D (1998) Odyssey 2 Technical Specs http://atarihq.com/danb/files/o2doc.pdf. Retrieved 25 June 2015
6. Bradley MM, Lang P (2000) Affective reactions to acoustic stimuli. Psychophysiology 37:204–215
7. Breinbjerg M (2005) The aesthetic experience of sound – staging of auditory spaces in 3D computer games. Aesthetics of Play. Bergen, Norway, October 14th–15th
8. Buchanan K (2013) You'll never guess how the dinosaur sounds in Jurassic Park were made. http://www.vulture.com/2013/04/how-the-dino-sounds-in-jurassic-park-were-made.html. Retrieved 07 July 2015
9. Cacioppo JT, Tassinary LG (1990) Principles of psychophysiology: physical, social, and inferential elements. Cambridge University Press, New York
10. Cantor D (1971) A computer program that accepts common musical notation. Comput Hum 6(2):103–109
11. Chanel G, Kierkels JJM, Soleymani M, Pun T (2009) Short-term emotion assessment in a recall paradigm. Int J Hum Comput Stud 67(8):607–662
12. Chang K, Kim G, Kim T (2007, August) Video game console audio: evolution and future trends. In Computer graphics, imaging and visualisation, 2007. CGIV'07
13. Cho J, Yi E, Cho G (2001) Physiological responses evoked by fabric sounds and related mechanical and acoustical properties. Text Res J 71(12):1068–1073
14. Cole T (2015) The tragedy of betrayal: how the design of Ico and Shadow of the Colossus elicits emotion
15. Collins K (2008) Game sound: an introduction to the history, theory, and practice of video game music and sound design. MIT Press
16. Collins K (2011) Making gamers cry: mirror neurons and embodied interaction with game sound. In Proceedings of the 6th audio mostly conference: a conference on interaction with sound, 39–46. ACM
17. Connor S (2010) Suspense-building music mimicks sounds of animals in distress. Independent Online. http://www.independent.co.uk/news/science/why-calls-of-the-wild-are-the-secret-of-a-good-horror-film-1982965.html. Retrieved 07 July 2015
18. Critchley HD, Mathias CJ, Dolan RJ (2002) Fear conditioning in humans: the influence of awareness and autonomic arousal on functional neuroanatomy. Neuron 33:653–663

19. Dekker A, Champion E (2007) Please biofeed the zombies: enhancing the gameplay and display of a horror game using biofeedback. Proceedings of DiGRA, 550–558
20. Dixon MJ, MacLaren V, Jarick M, Fugelsang JA, Harrigan KA (2013) The frustrating effects of just missing the jackpot: slot machine near-misses trigger large skin conductance responses, but no post-reinforcement pauses. J Gambl Stud 29(4):661–674
21. Drachen A, Nacke LE, Yannakakis G, Pedersen AL (2010) Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. 5th ACM SIGGRAPH Symposium on Video Games. 49–54. ACM
22. Duggan M (2008) Torque for teens. Cengage Learning
23. Ekman I (2008) Psychologically motivated techniques for emotional sound in computer games. Audio Mostly 2008 (Pitea, Sweden)
24. Fischer G, Grudin J, Lemke A et al (1992) Supporting indirect collaborative design with integrated knowledge-based design environments. Hum Comput Interact 7(3):281–314
25. Gilroy SW, Porteous J, Charles F, Cavazza MO (2012) Exploring passive user interaction for adaptive narratives. Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces. Lisbon, Portugal, 14th–17th February 2012, ACM, New York, pp 119–128
26. Gonzalez-Sanchez J, Chavez-Echeagaray ME, Atkinson R, Burleson W (2011) ABE: an agent based software architecture for a multimodal emotion recognition framework. In: Proceedings of the ninth working IEEE/IFIP conference on software architecture. pp 187–193
27. Graetz JM (1981) The origin of spacewar. Creat Comput 7(8):56–67
28. Grimshaw M, Garner T (2015) Sonic virtuality: sound as emergent perception. Oxford University Press, New York
29. Grimshaw M, Lindley CA, Nacke L (2008) Sound and immersion in the first-person shooter: mixed measurement of the player's sonic experience. Audio Mostly 2008, Piteå, Sweden
30. Hamalainen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounesmaa OV (1993) Magnetoencephalography: theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Physiol 65:413–497
31. Hamilton MF, Blackstock DT (1998) Nonlinear acoustics, vol 427. Academic, San Diego
32. Headlee K, Koziupa T, Siwiak D (2010) Sonic virtual reality game: How does your body sound? In: International conference on new interfaces for musical expression, Sydney, Australia, pp 15–18
33. Hermann T, Hunt A, Neuhoff (2011) The sonification handbook. Logos Verlag, Berlin
34. Huron DB (2006) Sweet anticipation: music and the psychology of expectation. MIT Press
35. Ismail F, Biedert R, Dengel A, Buscher G (2011) Emotional text tagging. http://gbuscher.com/publications/IsmailBiedert11_EmotionalTextTagging.pdf
36. Jennett C, Cox AL, Cairns P et al (2008) Measuring and defining the experience of immersion in games. Int J Hum Comput Stud 66(9):641–661
37. Jørgensen K (2006) On the functional aspects of computer game audio. Audio-Mostly 2006, Pitea, Sweden
38. Koelsch S, Kilches S, Steinbeis N, Schelinski S (2008) Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. PLoS ONE 3(7)
39. Kuikkaniemi K, Laitinen T, Turpeinen M, Saari T, Kosunen I, Ravaja N (2010) The influence of implicit and explicit biofeedback in first-person shooter games. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 859–868
40. Lang PJ, Bradley MM, Cuthbert BN, Patrick CJ (1993) Emotion and psychopathology: a startle probe analysis. Prog Exp Pers Psychopathol Res 16:163–199
41. Linzmayer O (1983) The voice of odyssey. Creat Comput Video Arcade Games 1(1):60
42. Lopes P, Liapis A, Yannakakis GN (2015) Sonancia: sonification of procedurally Generated Game Levels. http://www.ccgworkshop.org/wp-content/uploads/2015/06/CCGW2015_submission_4.pdf. Retrieved 09 July 2015
43. McAllister N (2011) Biometrics – the future of video games? Edge Online, http://www.edge-online.com/features/biometrics-future-videogames. Retrieved 11 Feb 2014
44. Mirza-Babaei P, Long S, Foley E, McAllister G (2011) Understanding the contribution of biometrics to games user research, DiGRA, pp 329–347

45. Moncrieff S, Venkatesh S, Dorai C (2001) Affect computing in film through sound energy dynamics International Multimedia Conference. Ninth ACM Int Conf Multimed 9:525–527
46. Murugappan M, Ramachandran N, Sazali Y (2009) Classification of human emotion from EEG using discrete wavelet transform. J Biomed Sci Eng 3(4):390–396
47. Nacke L, Ambinder M, Canossa A, Mandryk R, Stach T (2009) Game metrics and biometrics: the future of player experience research. Future Play, 2009
48. Nacke LE, Kalyn M, Lough C, Mandryk RL (2011) Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In: SIGCHI conference on human factors in computing systems. pp 103–112
49. Oculus VR (2016) www.oculus.com. Accessed 29 Sep 2016
50. Parker JR, Heerema J (2008) Audio interaction in computer mediated games. Int J Comput Games Technol 2008:1
51. Perron B (2004) Sign of a threat: the effects of warning systems in survival horror games. Cosign 2004 Proceedings. Art Academy, University of Split, pp 132–141
52. Plans E, Morelli D, Plans D (2015) AudioNode: prototypical affective modelling in experience-driven procedural music generation. http://www.ccgworkshop.org/wp-content/uploads/2015/06/CCGW2015_submission_7.pdf. Retrieved 09 July 2015
53. Poh MZ, Loddenkemper T, Swenson NC, Goyal S, Madsen JR, Picard RW (2010) Continuous monitoring of electrodermal activity during epileptic seizures using a wearable sensor. In: Conf Proc IEEE Eng Med Biol Soc. pp 4415–4418
54. Ravaja N (2004) Contributions of psychophysiology to media research: review and recommendations. Media Psychol 6:193–235
55. Sato W, Fujimura T, Suzuki N (2008) Enhanced facial EMG activity in response to dynamic facial expressions. Int J Psychophysiol 70(1):70–74
56. Shilling R, Zyda M, Wardynski EC (2002) Introducing emotion into military simulation and video game design America's Army: operations and VIRTE. GAME-ON
57. SPC700 APU Manual (1997) http://snesmusic.org/files/spc700_apu_manual.txt. Retrieved 03 July 2015
58. Steckler L (1975) TV Games Al 'I-Iome
59. Stony Brook University (2011) Tennis for two: history, http://www.stonybrook.edu/libspecial/videogames/tennis.html. Retrieved 02 July 2015
60. Sweet M (2014) Writing interactive music for video games: a composer's guide. Pearson Education
61. Tan ES (2000) Emotion, art, and the humanities. In: Lewis M, Haviland-Jones JM (eds) Handbook of emotions, 2nd edn. Guilford Press, New York, pp 116–134
62. Väisänen J, Väisänen O, Malmivuo J, Hyttinen J (2008) New method for analysing sensitivity distributions of electroencephalography measurements. Med Biol Eng Comput 46:101–108
63. Video Game Consile Library (2014) http://www.videogameconsolelibrary.com/pg70-bally.htm#page=specs. Retrieved 20 June 2015
64. Visch VT, Tan ES, Molenaar D (2010) The emotional and cognitive effect of immersion in film viewing. Cognit Emot 24(8):1439–1445
65. Winter D (2013) Atari PONG: the first steps, http://www.pong-story.com/atpong1.htm. Retrieved 01 July 2015
66. Woodward JD, Orlans NM, Higgins PT (2003) Biometrics: [identity assurance in the information age]. McGraw-Hill/Osborne, New York
67. Xu M, Chia L, Jin J (2005) Affective content analysis in comedy and horror videos by audio emotional event detection. In: IEEE International Conference on Multimedia and Expo
68. Yokota T, Fujimori B (1962) Impedence change of the skin during the galvanic skin reflex. Jpn J Physiol 12:200–209
69. Yoo MJ, Beak JW, Lee IK (2011) Creating musical expression using kinect. In: Proceedings of New Interfaces for Musical Expression. Oslo, Norway

Chapter 13
Emotional Appraisal Engines for Games

Joost Broekens, Eva Hudlicka, and Rafael Bidarra

Abstract Affective game engines could support game development by providing specialized emotion sensing (detection), emotion modeling, emotion expression and affective behavior generation explicitly tailored towards games. In this chapter we discuss the rationale for specialized *emotional appraisal engines* for games, analogous to having specialized physics engines. Such engines provide basic emotion modeling capabilities to generate emotions for Non Player Characters (NPCs), just like the *Havok* engine provides physics-related special purpose processing. In particular, such engines provide NPCs with an emotional state by simulating the emotional meaning of an event to an NPC in the context of the game's storyline, the NPC's personality, and relationships with other NPCs. We discuss why such engines are needed, present an example approach based on cognitive appraisal, and show how this appraisal engine has been integrated in a wide variety of architectures for controlling NPCs. We conclude with a discussion of novel gameplays possible by the more sophisticated emotion modeling enabled by an emotion appraisal engine.

Introduction

Emotions are arguably the most important element of gaming and game design [1–6]. Players experience a wide range of emotions during the entire gameplay process. This begins even before playing the game in the form of anticipation (hope) raised by advertisement or previous experiences and disappointment or confirmation of the hope raised by developer blogs about the expected game features. Then the process of ordering (unpacking) and installing involves a wide variety of affective experiences including frustration, eagerness, anger, relief, and concentration. The first actual contact with the game can involve feelings of awe and belonging (e.g. in

J. Broekens (✉) • R. Bidarra
Intelligent Systems Department, Delft University of Technology, Delft, The Netherlands
e-mail: Joost.broekens@gmail.com; R.Bidarra@tudelft.nl

E. Hudlicka
College of Computer Science, University of Massachusetts-Amherst & Psychometrix Associates, Amherst, MA, USA
e-mail: hudlicka@cs.umass.edu

the case of a well-designed sequel). Thus by the time we are actually playing the game, we have gone through an important process of emotional investment in the game. This creates a bond between the game and the player, sometimes so strong that this bond can have serious negative consequences for the player and his/her surroundings. On the other hand that same bond creates a willingness to invest time and effort in playing the game and generates unique positive experiences not possible with any other media.

All four of the core areas of affective computing can contribute to a more immersive and engaging gameplay experience: emotion sensing and recognition, emotion modeling, emotion-driven Non Player Character (NPC) behavior and expression of emotions by NPCs, as well as affective player modeling. Other chapters in this volume discuss emotion sensing and recognition in gaming as well as emotion-driven expression and NPC behavior, and affective content. This chapter focuses on emotion modeling proper, that is, the simulation of emotions for the Non Player Characters in a game. "Simulating emotions" refers here to the process of automatically determining when an NPC should express (or behave according to) a particular emotion and with what intensity. This includes two broad categories of underlying processing: *generation of NPC emotion* in response to some set of emotion-eliciting stimuli within the gameplay, and *modeling of the effects of these emotions* on the NPCs' internal processing (e.g., perception, decision-making, planning) and, ultimately, behavioral choices. In other words, we refer to computational modeling of emotion in the sense proposed by Hudlicka [7, 8]. In this chapter we focus on the first process: the generation of emotion in response to emotion-eliciting stimuli, called *appraisal*. Further, we focus on the "when to, what to, and how much to express", not on the "how to express and what to do"; i.e., in this chapter we are not concerned with rendering emotional expressions or generating particular behavior that should follow particular emotions (e.g., fear followed by the behavior of fleeing). These aspects of emotional processing fall within emotion expression processing and are addressed in other chapters in this volume.

NPCs in many existing games do possess emotional behavior (expression, actions). However, these emotions are typically scripted, event-triggered, or built-in within the storyline. As such, emotions in NPCs are more often considered a cosmetic feature, related to rendering and behavioral realism, rather than an integral component of the game mechanics that can influence gameplay. The lack of emotional complexity and flexibility in existing NPCs severely limits their affective realism, and, consequently, the realism and immersive potential of the resulting gameplay. To address this limitation, NPCs need to include deeper models of affective processing, as outlined above. Such model-based emotions can simulate plausible emotional reactions at appropriate moments during gameplay automatically, thereby increasing the believability of the non-player character [9, 10] by increasing the variability of NPC behavior [11] and enabling novel forms of gameplay [12–14]. For a more in-depth analysis of the rationale for including deeper models of emotions in NPCs, as well as existing approaches, see [3, 15, 16]. By "plausible emotional reaction" we mean that the emotion of the NPC should make sense to the player; i.e., the emotion should be *psychologically valid*, within

the context of the gameplay. Note that psychological validity does not imply that the emotion must be "normal", according to what normal individuals would feel in a particular setting. Rather, by a *psychologically valid model* we mean that *NPC emotions are simulated by a model that is based on what is known about human emotional appraisal*. This includes the possibility of generating unusual, abnormal or pathological characters, including plain evil characters, as these often fulfill important roles in a game's narrative.

In this chapter we discuss why specialized *emotional appraisal engines* for games are needed (section "Why Are Model-Based NPC Emotions Rare in Commercial Games?") and possible (sections "Emotional Appraisal Engines as Plug-in Modules", "Integrating Emotional Appraisal Engines with NPC Control", and "Appraisal Engines Enable Novel Gameplays and Genres"). The term 'appraisal engine' refers to specialized game engines that support modeling of emotions in NPCs, in a manner that does not require a commitment to a particular NPC architecture. In other words, emotional appraisal engines provide a modular approach to augmenting NPCs with emotion, analogous to a plug-in.

The term 'appraisal engine' contrasts with the term 'affective game engine' [11]. As originally envisioned, affective game engines would provide a broad range of tools to support the development of affective and affect-adaptive games by providing functionalities to facilitate implementing all four of the core areas of affective computing: recognition of player emotions, emotion modeling in NPCs (including both emotion generation and modeling of emotion effects on the NPCs internal processing and behavior), expression of emotions by NPCs, and affective player modeling. Currently, no game engine exists that provides all of these functionalities.

While many computational models of emotion are available [17–19] that could be used in games and experimental games exist that use some of these models [13, 20], commercial games have not incorporated model-based emotions (with some exceptions, see [3, 15]). We believe this is due to several reasons including: challenges associated with integrating cognitive appraisal models with existing NPC AI; lack of standard testing and development tools, resulting in perceived high risk and development costs; the need to understand the mechanisms mediating emotion elicitation; lack of examples of novel gameplays enabled by emotional NPCs; and, the conviction that gamers don't ask for emotional NPCs.

Support for these arguments can be found in a recent pilot study in which we investigated game designers' perspectives on emotional characters. Results from this study will be discussed in more detail in section "Why Are Model-Based NPC Emotions Rare in Commercial Games?". To show how these causes can be mitigated, we discuss recent projects in the area of development and integration of emotional appraisal engines. In section "Emotional Appraisal Engines as Plug-in Modules", we briefly present a recent approach towards developing an emotional appraisal engine [15]. We show how this approach addresses several of the issues outlined above, specifically: how it eliminates the dependency between existing NPC AI and the mechanisms required for dynamic emotion simulation; how it reduces the need to understand the emotion elicitation process; and how it enables control over the emotional behavior of NPCs. In section "Integrating Emotional

Appraisal Engines with NPC Control" we show that proper encapsulation of emotional appraisal enables integration with widely different NPC control mechanisms. In section "Appraisal Engines Enable Novel Gameplays and Genres", we present examples of novel gameplays enabled by NPCs augmented with model-based emotions. Some of these novel types of gameplays can be considered novel genres [14] that go beyond the traditional emotional deepening of the relationship between the player and the NPC [5].

Why Are Model-Based NPC Emotions Rare in Commercial Games?

We believe there are several reasons why NPCs with emotions based on a computational model are rare in commercial games. These reasons are technical, conceptual and financial (see list below). These reasons can also been seen as a list of concerns regarding the feasibility of implementing model-based emotions in NPCs, and thus provide a set of requirements for affective game engines in general, and appraisal engines in particular.

1. Most emotion models are dependent on particular NPC AI: *complexity and modularization* (*technical*)
2. Lack of tools for design, development, and testing: *dev. support* (*technical*)
3. Complexity of modeling emotion elicitation: *complexity of emotions* (*conceptual*)
4. Lack of emotion-enabled gameplay innovation: *gameplay* (*conceptual*)
5. Players do not demand emotional NPCs: *market demand* (*financial*)
6. Waiting for a big game publisher to bite the bullet: *financial risk* (*financial*)

From a technical point of view, emotion modeling is rarely adapted to the game development process. Even though emotion modeling, especially based on the Ortony, Clore and Collins model (OCC) [21], has been used [22] and analyzed extensively [23–25] in research, none of these approaches have led to emotion simulation in games by means of computational modeling.

A major reason is that emotion modeling in research usually adopts a particular type of Artificial Intelligence NPC architecture, with its associated representational and reasoning mechanisms (e.g., fuzzy logic, BDI-agents), as basis for the computational model of emotion. However, just as a game developer does not want its destroyable assets to follow a particular morphology when they use a special purpose physics engine to simulate destruction, game developers also do not want to use a particular type of AI for their agents when simulating emotions. These aspects of game design need to be, and can be, decoupled [15]. *A successful appraisal engine thus needs to be an AI-independent module*, *with a clear Application Programming Interface* (*API*).

A second major reason is that developing a game involves a great deal of design and testing and the tools supporting these activities are lacking. It is also unclear how one should go about debugging emotions of NPCs in the first place. What kind of visualizations are needed? What kind of aggregation over NPCs and over time is needed to quickly get an idea of what happens emotionally to a set of NPCs? What is the exact behavior of a particular computational model in different settings? Just as a physics engine has predictable behavior with regards to physics (at least to a certain extent), an emotion simulation engine also needs some way of giving the game designer a sense of control over the resulting NPC behaviors. The difficulties associated with testing of emotional behavior of NPCs have also been confirmed with our pilot study (Fig. 13.1). *A successful appraisal engine needs to incorporate design and testing tools that give the game developer insight into the range of*

	Test Value = 3					
Statement	t	df	Sig. (2-tailed)	Sig. (1-tailed)	Mean	(Dis)Agree
I am interested in developing games with emotional NPCs	**2.420**	**12**	**0.03**	**0.02**	**3.7**	**Agree**
I am familiar with the concept of emotional NPCs and how they generate emotions.	.762	12	0.46	0.23	3.2	
I see the purpose of emotional NPCs in games.	**3.825**	**12**	**0.00**	**0.00**	**3.8**	**Agree**
I understand how emotional NPCs can enhance the gaming experience for players.	**2.920**	**12**	**0.01**	**0.01**	**3.7**	**Agree**
Apart from emotion expression, emotions in NPCs do not add value to gameplay.	-.485	12	0.64	0.32	2.8	
Players demand emotional NPCs.	**-3.317**	**10**	**0.01**	**0.00**	**2.0**	**Disagree**
Players will not notice emotional capabilities in NPCs	.562	12	0.58	0.29	3.2	
Players do not care if NPCs have emotional capabilities.	-1.148	12	0.27	0.14	2.8	
Publishers are not putting emotional NPCs on their feature lists.	1.600	12	0.14	0.07	3.6	
I would use emotional NPCs if my competitors start using them	-1.328	12	0.21	0.10	2.6	
I understand the cost associated with adding emotional NPCs to my games.	0.000	12	1.00	0.50	3.0	
The initial investment (licenses, etc.) to add emotional NPCs to my games is steep.	1.477	12	0.17	0.08	3.3	
The extra game assets required for emotional NPCs are expensive.	**2.856**	**12**	**0.01**	**0.01**	**3.8**	**Agree**
I understand how to develop emotional NPCs.	.210	12	0.84	0.42	3.1	
Incorporating emotional NPCs to the game design is complex.	**3.207**	**12**	**0.01**	**0.00**	**3.9**	**Agree**
Programmers do not have the necessary knowledge to develop emotional NPCs.	-.485	12	0.64	0.32	2.8	
Games with emotional NPCs are hard to test.	1.443	12	0.17	0.09	3.4	
It is hard to test the behavior of emotional NPCs.	**2.889**	**12**	**0.01**	**0.01**	**3.6**	**Agree**
I am confident in releasing games where the NPCs have autonomous emotional behavior.	-.457	12	0.66	0.33	2.8	
My confidence in emotional NPCs behaving appropriately would increase if a specialized testing suite was available.	-.519	12	0.61	0.31	2.8	
If standard development methodology was available I would start adding emotions to my NPCs.	-.457	12	0.66	0.33	2.8	

Fig. 13.1 (Dis-) agreement with statements about emotional NPCs based on 13 subjects (5 international, 8 Dutch) from different game development companies. Subjects rated agreement on a 5-point scale. (Dis-) agreement (*bold rows*) was decided based on statistical significance of a 1-tailed T-test (test value = 3)

emotions, *the intensity of emotions*, *and the causes for emotions at different levels of aggregation* (*individual NPCs*, *groups of NPCs*, *and even game world areas*).

For non-specialists, it is difficult to understand what emotion modeling really is, and what it can bring to the gameplay. For example, if one assumes that NPC emotions serve to evoke deep player-NPC interaction (an opinion explicitly voiced in our pilot experiment by one of the participants, as well as in [5]), then of course modeled NPC emotions are limited to adventure and RPG-like genres. However, this is not the case. Even simple puzzle games can benefit from, or even completely revolve around, emotion simulation ([14], and section "Appraisal Engines Enable Novel Gameplays and Genres"). If one assumes that modeled emotions are always psychologically plausible, and that this means that the emotions are those of healthy and sane individuals, this excludes emotion modeling for antagonists, as these characters usually are mean, deranged, mad, or otherwise not normal. It also excludes the possibility of developing serious games for psychotherapy, which may require the modeling of abnormal affective reactions and behavior motivated by emotion dysregulation. However, this is also not the case. Computational models of emotion are based on an understanding of affective processes based on psychological emotion theories, and these theories can also be applied to simulating "bad guys" in entertainment games, and pathological affective behavior of NPCs in therapeutic games. Psychological plausibility simply means that the emotions resulting from the model are predictable from a psychological point of view, not that they are "normal" or appropriate to experience at a particular moment in time. Physics engines can model realistic destruction just as easily as realistic movement; the laws of physics don't change and the laws of emotion don't change either. If one assumes that it is difficult to add emotional NPCs to the game design (see pilot results), that emotion simulation means that a full range of affective characteristics must be included (mood, personality, etc.), or that expensive additional game assets are needed for emotions to bring a noticeable difference to the player (pilot study), then one puts up unnecessary walls. Even simple emotions, generated by a rudimentary appraisal engine, can provide interesting game mechanics ([14], and section "Appraisal Engines Enable Novel Gameplays and Genres"). *An appraisal engine for games should hide the complexity of emotion elicitation*, *and provide easy to use functionality and flexible control for simple and complex cases of emotional NPCs*.

Putting up such walls limits the designers' creativity. A telling indication that this occurs is that game developers state that even though *they* see how emotional NPCs could be used in games, players don't demand emotional NPCs (pilot study). For these walls to crumble, different emotion-based gameplays need to be developed and games need to hit the charts. In fact this has happened already with one emotional game genre in the form of the *The Sims* series by Electronic Arts. The Sims' emotions illustrate how emotion models can be used to add realism and fun to the characters in a management/simulation type game. Although successful, this is only one way in which emotions can impact gameplay. Perhaps there is a perceived risk for other developers in adding emotion modeling to NPCs because this means entering the domain of another developer. This risk is real. Apart from

the effort to add novel technology that needs to pay off at some point, a company might risk "collateral damage" to its reputation for "trying but failing" to do a better job. To understand the potential of adding emotions to NPCs, experimenting with novel game design and novel gameplays in smaller and simpler games is essential. Developing a computational model of emotion that "runs" is not sufficient. *A successful emotion engine for games should make explicit how the engine facilitates novel gameplay*, *otherwise the engine does not have perceived value*.

In the remainder of this chapter we discuss how some of the concerns about model-based emotions for NPCs can be addressed. We do this by discussing our own work, because we know this work best. We do not claim we are the only ones trying to address these issues. In our view, these concerns revolve around three aspects of emotion modeling: the need for a black-box, easy to use emotional appraisal engine; the need for easy integration in different NPC architectures; and the need for seeding creativity with novel emotion-enabled game genres. A black-box emotional appraisal model that is easily integrated in different settings is needed to resolve the problem of AI dependency. Novel game genres are needed to explore whether there is a market for games with emotional NPCs. If potential players are not presented with interesting games with novel gameplays, then demand for such games will never arise. Recall that no players asked for Pac-Man either.

Emotional Appraisal Engines as Plug-in Modules

To facilitate the development of emotional NPCs, an emotional appraisal engine needs to support the development of both simple and complex emotional NPCs, as well as hide the complexity of the emotion generation process, and simulate the emotions in an AI-independent way. In addition it should be high performing and scale efficiently [26]. In other words, the emotion engine should function analogously to a physics engine: it should provide emotion simulation functionality with a clearly defined API that would not depend on the AI used for controlling the NPCs in the game, as this would limit the engine's capacity to operate in a wide variety of games. To demonstrate that this is feasible, we have developed an emotion appraisal engine, GAMYGDALA [15]. Below we briefly describe how the basic assumptions of this appraisal engine help achieve the aforementioned goals.

GAMYGDALA (available in Java, Javascript, and C#) provides three main functions: *emotion generation* via *cognitive appraisal* based on the OCC model; *dynamic relationships among NPCs*, based on a simple like/dislike scheme; and NPC *affective dynamics modeling* including integration of emotional appraisal over time, intensity and decay, and a translation of the categorical emotional state to a the dimensional abstraction defined by the dimensions of Pleasure, Arousal, Dominance (PAD).

To use GAMYGDALA, a game developer first defines the NPC's set of goals (these can include achievement, maintenance and avoidance goals, and can change over time). The developer then decides how particular game events impact these

goals, and sends these annotated events to the appraisal engine for processing. These specifications are sufficient to elicit emotions in the NPC. In addition, the developer can configure relationships among NPCs (relationships also develop as a result of emotional appraisal). The NPC relationships then form the basis for social emotions such as feeling resentment (negative relation towards another and a positive event happening to that other) or feeling happy-for another agent (positive relation and positive event). Goals and events need to be specified only for the event appraisal necessary for emotion generation, and need *not* be related to the game AI.

GAMYGDALA is game-AI independent because it only performs a "black-box" appraisal function. It defines the minimum interface needed to implement (a subset) of the OCC model and addresses a baseline level of affective dynamics. To demonstrate GAMYGDALA's AI independent nature, we discuss in section "Integrating Emotional Appraisal Engines with NPC Control" its integration with a cognitive agent programming language [27], a system for defining semantic game worlds [28], a narrative generation engine [29], and with *Phaser* (a Javascript game engine). In all cases the same interface was used based on the annotation of events and the definition of agent goals, as outlined above.

It is possible to simulate emotions for a diverse range of NPCs, especially simpler cases of NPCs, where the set of goals for one NPC is limited [14]. More complex cases, such as when an agent has a large set of goals that include achievement and avoidance goals, when it is important that the agent interpret the event in relation to previous events, or when an event acts on two different goals in opposing directions (e.g., where one is an achievement and the other an avoidance goal), need to managed by the game developer. "Managed" in this case means that the developer needs to decide which event to appraise in relation to which goals, because GAMYGDALA is simply a black-box appraisal engine, and does not implement any reasoning, memory or attentional processing itself. While this limits GAMYGDALA's functionality, implementing these functionalities in an appraisal engine would necessitate making assumptions about game AI and NPC behavior generation, and would violate GAMYGDALA's objective of maintaining independence from the existing NPC AI architecture.

Integrating Emotional Appraisal Engines with NPC Control

In this section we demonstrate the notion of emotional appraisal engine independence from the NPC AI by illustrating how an emotional appraisal engine can be used with widely different ways of controlling NPC behavior. In particular, we show how we used GAMYGDALA to model the emotions of NPCs developed with the *Phaser* game engine framework, with the cognitive agent programming language GOAL [27], a system for generating semantic worlds and crowds called Entika [28], and with narrative generation [29]. This section thus serves as proof for the claim that black-box appraisal is an approach that enables reusability of the emotional appraisal functionality and can encapsulate emotion modeling complexity.

Simulating Emotions in Semantic Worlds

Creating a virtual world where interaction with almost every object is possible poses many difficult challenges that are far from being solved, including the complexity of maintaining all possible interactions with, and among, the game entities [30]. In particular, to enable the objects with which the player interacts to have an emotional impact on an NPC requires defining some meaning, or *semantics*, for the objects and for those interactions. One approach aimed at solving this problem is *semantic game worlds* [28]. These worlds are designed by defining and choosing the entities populating them from among specific classes, each with their own unique properties, such as attributes, roles and services [31]. With this approach, the objects within the game world themselves carry the information of what they are actually useful for, or able to. It is therefore possible for NPCs to query the game world and discover usable objects for their purposes. In order to facilitate the creation of semantic worlds, a framework called *Entika* was developed [28], that supports a simple and intuitive definition of semantics, promotes re-usability and facilitates object behavior customization. More recently, this framework has been applied to the specification and simulation of the motion behavior of a crowd of agents. For this, a semantic crowd editor was developed aimed at defining crowd templates in a portable way, allowing their reuse for virtually any environment in which the available objects are spontaneously used by other agents in a meaningful manner. This is achieved by having each agent query the environment in order to find whatever objects are deemed suitable to fulfill its goals [32].

With this basis, implementing NPC emotions in a game world involves supporting and integrating two essential elements: semantics and emotions. The semantics component involves creating a semantic game world, populated with entities (living or otherwise) with clearly defined semantics. The emotions component involves analyzing how (un-)desirable the interaction with objects an actors is for the goals of each character, and generating the corresponding emotion(s) accordingly. Semantics in the context of emotion simulation can be rephrased as *the information needed to perform emotional appraisal*. This is the approach behind the integration of Entika and GAMYGDALA.

Goals are instrumental to the simulation of emotion, and, because every NPC in a game should have some goal(s), defining goals for an NPC is a natural feature of a game world with living entities. Further, whatever happens to the NPC can be seen as an *event*, which consists of an action performed at a given moment in time, involving one or more entities. In addition, events can be associated with several other attributes, so that it becomes possible to specify the goals influenced by the event, as well as to indicate whether this influence is positive or negative. The integration boils down to the following: Entika is used to specify events and NPC goals, and GAMYGDALA processes these events and appraises them in the relation of NPC goals at runtime. The flexibility of Entika is leveraged to defining semantic game worlds, while the emotional appraisal functionality is provided by GAMYGDALA to emotionally interpret what happens in the world.

For the configuration of the appraisal engine, various parameters are involved that are provided by Entika such as *belief likelihood*, modeled as the credibility of an event happening; *utility*, specified for each goal of an NPC, indicating the degree (positive or negative) to which the entity wants a given goal to be fulfilled; and *goal compatibility*, i.e., the extent (either positive or negative) to which an event influences the likelihood of a goal being achieved.

The coupling between Entika and GAMYGDALA demonstrates two concepts. First, that appraisal can be easily embedded in a semantic world. For this, abstractions that are relevant to an agent's emotional state are very naturally developed using semantics. Notions such as goals, beliefs and agents can be semantically tied to one another in a way that is generic, scalable, intuitive and independent of implementation platform. Second, that it is feasible to define virtual environments based on semantics, which yield plausible emotional states for NPCs by using a generic emotional appraisal engine. For the full report of this study see [33].

Simulating Emotions in Phaser, a Javascript Game Engine

Phaser is a Javascript-based game engine. A Javascript plug-in for Phaser has been developed as a wrapper around GAMYGDALA's public API. Using this plug-in one can configure the appraisal engine to appraise events in the following way. For every NPC that needs emotions, the game developers generate an agent entity in the appraisal engine. The developer then creates goals for each agent, including a goal utility ([-1, 1]). Finally, if needed, the developer creates a positive or negative relationship ([1–, 1]) between NPCs, using a single *createRelation* method. When an event needs appraising, the developer calls the *appraise* method, and defines how the event impacts particular goals. GAMYGDALA then determines which agents are affected and what emotions result from the appraisal, as well as how the NPCs' relationships are affected. This simple API allows the generation of emotional NPCs in a way that shields the developer from the complexity of appraisal itself, and is another example of how black-box appraisal can facilitate the integration of emotions in a different type of NPC development environment. Of course, the integration approach is the same as for the integration with Entika: define agents, define goals and define how the events impact the goals.

Simulating Emotions in Cognitive Agent Programming

A very different setting for the controlling of NPC behavior is cognitive agent programming. Here an agent (or NPC) is controlled using cognitive reasoning in the form of rules that trigger actions based on preconditions that need to be met in order for the rules to fire. In the agent programming language GOAL [27], such agent behavior is specified by a programmer who defines rules, goals and domain

knowledge. An agent has a current mental state, representing the state of affairs, and the agent reasons using a rule set about what to do next based on this current mental state. When a rule fires, it triggers an action. Examples of actions are movements, firing bullets, or anything else that can be implemented by the physical embodiment of the agent in its environment. GOAL has been used to control game bots in, for example, Unreal Tournament [34]. GAMYGDALA has been added to GOAL in two different ways: as an integration in the reasoning cycle, and, as a plug-in emotional appraisal module. Here we only explain the integration as a plug-in.

The appraisal engine GOAL plugin works in a similar manner as the integration with Entika. At the creation of a GOAL agent instance (after launching a world with its agents), GOAL launches the plug-in and creates a GAMYGDALA agent instance. At design time, the agent programmer defines GOAL program rules in which the appraisal plug-in is instructed to add and remove goals, as well as appraise events. These instructions are implemented as built-in GOAL actions with similar arguments as described for the Entika and Phaser integration above. These actions call the appraisal's plug-in functionality, and after each appraisal the resulting emotions are added to the agent's mental state as belief predicates, e.g., *emotion*(*happy*, *0.8*). Again, this integration takes place in a very different development environment, but the integration approach is the same.

Simulating Emotions in Narrative Generation

A different approach to integrate appraisal was used in the affective storyteller, a system that generates stories with simulated actor emotions based on the events that happen to the actors in the story [29]. In this system, GAMYGDALA is integrated into the story generation process. The story generation is driven by actors taking actions based on their goals and the current state of the story. Actors and goals are also configured in the appraisal engine, and whenever an event occurs in the story it is automatically annotated and sent to the appraisal engine for processing. The automatic annotation is possible because event likelihood and goal congruence are derived from the story at runtime. Goals are annotated with a utility at design time. This means that once the story domain has been set and the actors are configured, the appraisal engine will fully automatically generate emotions. This is a different example of an integration, where the integration is more closely coupled to the AI than in the previous examples. Here, the appraisal engine gets all of its information, except the goal utilities, from the AI that generates the story.

Appraisal Engines Enable Novel Gameplays and Genres

The discussion above outlined the motivation for, and benefits of, augmenting NPCs with more sophisticated models of emotions and providing a specialized emotional appraisal engine to facilitate this modeling. The appraisal engine discussed here

focuses on a subset of the emotion modeling: emotion generation via cognitive appraisal. (A full-fledged emotion engine would also provide the tools to facilitate incorporating the effects of emotions on the NPCs behavior and internal information processing, including perception, decision-making and planning.)

Below we discuss how even relatively simple emotional appraisal engines, such as those discussed in this chapter, can enhance gameplay. We also highlight novel gameplays, even novel game genres, made possible by dynamic models of affective processing in NPCs. We provide examples of such novel gameplays for several of the more popular game genres, and highlight in particular the benefits of explicit emotion modeling in NPCs in serious games.

Action-Adventure Games (e.g., Legend of Zelda, The Witcher)

These games involve a combination of exploration and puzzle/problem solving, where the puzzles/problems range in complexity from simple, concrete tasks (unlock door to retrieve object) to complex interactions with NPCs. Dynamic generation of even the basic emotions in the NPCs would enable them to display more variability in behavior, thereby providing both increasing affective realism, and "surprises" during the gameplay. Models of more complex social emotions would also provide the opportunity for creating sophisticated 'social puzzles', where the player's progress in the game would necessitate inducing a particular emotion in the NPC, thus shifting the realm of the puzzles from manipulating physical objects to creating and managing complex social interactions. Games like *Crusader Kings II* already take up such an approach, where management of relationships is an important aspect in growing the player's medieval empire.

What an appraisal engine can bring to this is the straightforward way of modeling how NPCs react to attempts to manipulate their emotional state. This would allow the game developers to embed this enhancement in the gameplay, even in games where emotion is not a key focus. It would also allow the generation of "emotional puzzles" for example as a proper mini game. See [14] for an implemented example of such an emotional puzzle game.

Fighting and First-Person Shooter Games (e.g., Mortal Combat, Doom)

Both of these genres focus on direct combat with and/or killing of the opponent. The player typically views the opponent directly, although some games provide a third-person view of the gameplay. The affective realism of the NPCs in this genre is limited to rudimentary expressions of aggression, and, less frequently, fear. In existing games these emotions are typically scripted, and little variability or nuanced NPC behavior is possible. Games augmented with model-driven emotion generation

would enable the NPCs to vary their emotional reactions to the evolving context, both in terms of which emotion is displayed when, and the intensity of that emotion; again, providing less predictable behavior and thus more engaging gameplay.

While the NPCs' display of aggression and rudimentary fear may be adequate for most FPS players, one can imagine an interesting evolution of this genre, where the opponent NPCs could display more complex dynamics of these emotions, and include additional emotions, such as sadness, guilt, envy, happiness or pride. Display of this broader set of emotions would not only enrich the player experience, but could also 'humanize' the typical FPS gameplay. Also, such social emotions could impact NPC capabilities later in the game (e.g., guilty NPCs fight less aggressively or adopt different tactics). While this might not be the experience sought out by most current FPS players, such augmentation could be a welcome development in this genre and might mitigate the violence de-sensitization effects that are attributed to FPS games.

In general, NPC opponents could adopt different strategies depending on their emotional state. For example, a fearful NPC opponent that is immune to particular player moves, because fear makes the NPC react more quickly and move in a more random manner. Beating the fearful NPC would mean chasing it until exhaustion, while beating an angry NPC could be done with a few well-placed special blows. The player could manipulate the emotional state of the NPC before or during the combat scene, by manipulating the game objects or via interaction with other NPCs. Such gameplay would be easily supported by an appraisal engine and would require relatively simple additional game assets to display the NPC emotional state to the player.

Real-Time Strategy Games (RTS) (e.g., Age of Empires, Warcraft)

RTS games involve a competition with another player or NPC for some set of resources. It is easy to see how increased affective sophistication of the NPCs would enhance the gameplay. As in the action adventure games above, not only would the element of surprise, enabled by model-driven emotion generation, provide a more engaging experience, but the possibility of inducing distinct emotions in the NPCs to motivate them to engage, or not engage, in specific behavior would create an entirely distinct domain for player behavior: the social domain. In contrast to the current RTS physical domains, where physical objects or resources are manipulated, the addition of a social domain, enabled by the NPCs affective realism, would allow the players to 'manipulate' and control social resources, such as the NPCs' good or ill will. This capability would create much more believable gameplays, reflecting more accurately the reality of political and military strategies this genre aims to emulate. This is in line with, for example, *Crusader Kings II* (a turn-based strategy game), where relationships play an important role in how the political and military struggles unfold over the different generations.

Role-Playing Games (RPGs) (e.g., Baldur's Gate, Skyrim)

RPGs are fantasy-adventure games where the player(s) engage in a complex series of tasks to achieve the ultimate game goal (e.g., save humanity from a plague, defeat an evil cult, etc.). Novel gameplays enabled by emotion modeling in NPCs include those outlined above for the real-time strategy games. However, the realm of social interaction and the possibility of achieving affective and social tasks is even more important in RPGs, since the player has the opportunity to develop deep interactions with the NPCs over the course of playing through different quests [5]. In addition, a variety of emotional state manipulation can be implemented and quests can be (de)activated based on the emotion of particular NPCs, thus making emotional manipulation a key element of the quest progression. The same applies to simulation games such as The Sims.

Arcade and Platform Games (Pac-Man, Mario Bros)

Even basic emotion simulation can enhance arcade-like gameplay. Behavior of "baddies" in the game world can be made dependent on their emotional state. This is analogous to the influence of emotions on NPCs in the FPS and Fighting genre. For example, emotion-augmented baddies might grow to hate or like you depending on what items you gather in a level. Levels could be built around the idea that the player has to manipulate the emotions of NPCs in such a way that certain baddies become a threat while others do not. An example of such gameplay is presented in [14] in the form of a game called *friend or foe*.

Serious Games

Currently, games are used primarily for entertainment purposes. Increasingly however, games are being adapted for instructional, training and therapeutic purposes. In fact, serious gaming is the fastest growing segment of the game industry. All of the above genres can be adapted to serious gaming and the increasing affective sophistication of the NPCs, facilitated by affective game engines, is even more relevant in serious gaming contexts, particularly in therapeutic serious games, including games used to augment psychotherapy.

   The objective of serious games in psychotherapy is to support and augment face-to-face therapy, by providing opportunities to both experience problematic situations and to practice new behavior and coping strategies [35, 36]. The immersive quality of games, particularly games that use elements of virtual reality (e.g., head-mounted displays) provide a unique means of creating customized physical and social situations that induce undesirable behavior, and the opportunities to develop, *in*

virtuo, strategies to cope with these situations. The simulation of psychologically plausible emotions is especially relevant for simulation of complex social situations.

An emotion game engine that would support the development of these functionalities in a manner that is verified with respect to psychological plausibility would be especially useful for developing serious games, because it would enable the game designers and developers to focus on the core of the training or treatment, while the implementation of the psychologically plausible NPC emotions would be provided by the emotion simulation module.

Final Remarks

In this chapter we have motivated the need for specialized emotional appraisal engines for games, to facilitate the modeling of dynamic and more complex affective behavior in NPCs. We focused on modeling emotion generation via cognitive appraisal and presented evidence that when appraisal is approached in a black box fashion, where the appraisal process is decoupled from the logic controlling the NPCs, it is possible to embed the appraisal process as a plug-in in many different NPC architectures. This chapter is limited to a discussion of emotion generation models, and focused on discussion of approaches that would facilitate the development of model-driven emotions for NPCs. The effects of emotions on the NPCs internal processing and behavioral choices, also part of emotion modeling, have not been addressed.

Currently, the state of the art in emotion modeling literature is not advanced enough to have standard and agreed upon models of how emotion influences agent behavior. (See [37, 38] for a discussion of existing efforts and challenges in this area.) Future emotion engines for games should also include explicit models of the effects of emotions on the NPCs internal information processing (e.g., perception, decision-making, planning), as well as behavior choices. These models could then be used by game developers to bias the NPCs' reactions in the game environment, including the social environments; that is, the NPCs' reactions to the behavior of the player and other NPCs. As is the case with models of emotion generation, here it would also be important to hide complexity of these models. Just as a black-box appraisal plug-in, a black-box emotion effects plug-in should also be independent of the AI used to control the NPC, and this is perhaps an even greater challenge. Existing efforts in this area provide the basis for facilitating this type of modeling in emotion game engines. For example, the MAMID modeling methodology [39, 40] provides a means of encoding the effects of a broad range of factors, including emotions and personality traits, on internal processing. The parameters control the speed, capacity and biasing within individual architecture modules, and facilitate not only the modeling of the distinct effects of emotions on NPC behavior, but also the rapid construction of a wide variety of distinct NPC personalities. These capabilities would provide additional tools for the game developer to facilitate the construction of increasingly affectively realistic NPCs.

Finally, an aspect we have not touched upon is testing and debugging emotional appraisal. Insight into how one can debug and test emotional models is needed to manage the risks involved in having autonomously behaving entities in a game. This is of specific importance in large, procedurally generated worlds, where game designers set constraints for the game generation but do not have full control over how the game world and its inhabiting NPCs are configured.

Notwithstanding these limitations, we have shown the potential of emotional appraisal engines to transform the development of affective games. Model-based emotion generation for NPC can take the use of emotions in game design to the next level, and can be a basis for many novel gameplays.

References

1. Schell J (2008) The art of game design: a book of lenses. Morgan Kaufman Publishers, Amsterdam
2. Karpouzis K, Paiva A, Isbister K, Yannakakis GN (2014) Guest editorial: emotion in games. IEEE Trans Affect Comput 5(1):1–2
3. Yannakakis GN, Paiva A (2014) Emotion in games. In: Calvo RA, D'Mello S, Gratch J, Kappas A (eds) Handbook on affective computing. Oxford University Press, New York, pp 459–471
4. Hudlicka E, Broekens J (2009) Foundations for modelling emotions in game characters: modelling emotion effects on cognition. In: Affective computing and intelligent interaction and workshops, 2009. 3rd international conference on ACII 2009, pp 1–6
5. Freeman D (2004) Creating emotion in games: the craft and art of emotioneering™. Comput Entertain (CIE) 2(3):15–15
6. Hudlicka E (2008) Affective computing for game design. In: Proceedings of the 4th Intl. North American Conference on Intelligent Games and Simulation, pp 5–12
7. Hudlicka E (2003) To feel or not to feel: the role of affect in human-computer interaction. Int J Hum-Comput Stud 59(1–2):1–32
8. Hudlicka E (2008) What are we modeling when we model emotion? In: AAAI Spring symposium 2008
9. Reilly WS, Bates J (1992) Building emotional agents. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA
10. Bidarra R, Schaap R, Goossens K (2010) Growing on the inside: soulful characters for video games. In: Proceedings of the IEEE conference on computational intelligence and games
11. Hudlicka E (2009) Affective game engines: motivation and requirements. In: Proceedings of the 4th international conference on foundations of digital games. ACM, Orlando, pp 299–306
12. McCoy J, Treanor M, Samuel B, Reed AA, Mateas M, Wardrip-Fruin N (2014) Social story worlds with comme il faut. Comput Intell AI Games IEEE Trans on 6(2):97–112
13. McCoy J, Treanor M, Samuel B, Mateas M, Wardrip-Fruin N (2011) Prom Week: social physics as gameplay. In: Proceedings of the 6th international conference on foundations of digital games. ACM
14. Broekens J (2015) Emotion engines for games in practice: two case studies using Gamygdala. In: Gratch J, Schuller B (eds) 2015 international conference on affective computing and intelligent interaction (ACII), Xi'an, pp 790–791
15. Popescu A, Broekens J, Someren MV (2014) GAMYGDALA: an emotion engine for games. IEEE Trans Affect Comput 5(1):32–44
16. Hudlicka E (2011) Guidelines for designing computational models of emotions. Int J Synth Emotions (IJSE) 2(1):26–79

17. Marsella SC, Gratch J (2009) EMA: a process model of appraisal dynamics. Cogn Syst Res 10(1):70–90
18. Aylett R, Vala M, Sequeira P, Paiva A (2007) In: Cavazza M, Donikian S (eds) FearNot! – an emergent narrative approach to virtual dramas for anti-bullying education, in virtual storytelling. Using virtual reality technologies for storytelling. Springer, Berlin, pp 202–205
19. Ochs M, Sabouret N, Corruble V (2009) Simulation of the dynamics of nonplayer characters' emotions and social relations in games. Comput Intell AI Games IEEE Trans on 1(4):281–297
20. Mateas M, Stern A (2003) Façade: an experiment in building a fully-realized interactive drama. In: Game developers conference
21. Ortony A, Clore GL, Collins A (1988) The cognitive structure of emotions. Cambridge University Press, New York
22. Marsella S, Gratch J, Petta P (2010) Computational models of emotion. In: Scherer KR, Bänziger T, Roesch E (eds) A blueprint for affective computing. Oxford University Press, New York, pp 21–45
23. Bartneck C (2002) Integrating the occ model of emotions in embodied characters. In: Proceedings of the workshop on virtual conversational characters, pp 39–48
24. Steunebrink BR, Dastani M,, Meyer J.-JC (2008) A formal model of emotions: integrating qualitative and quatitative aspects. In: Proceedings of the European Confernce on Artifical Intelligence (ECAI '08). IOS Press, pp 256–260
25. Steunebrink BR, Dastani M, Meyer J.-J.C (2007) A Logic of emotions for intelligent agents, In: Proceedings of the 22th national conference on artificial intelligence (AAAI '07). AAAI Press, pp 142–147
26. Broekens J, DeGroot D (2004) Scalable and flexibel appraisal models for virtual agents. In: Mehdi Q, Gough N (ed) Proceedings of the fifth game-on international conference, pp 208–215
27. Hindriks K, Meyer JJ (2009) Toward a programming theory for rational agents. Auton Agents Multi-Agent Syst 19(1):4–29
28. Kessing J, Tutenel T, Bidarra R (2012) Designing semantic game worlds. In: Proceedings of the third workshop on Procedural Content Generation in Games. ACM, Raleigh, pp 1–9
29. Kaptein F, Broekens J (2015) The affective storyteller: using character emotion to influence narrative generation. In: Brinkman W-P, Broekens J, Heylen D (eds) Intelligent Virtual Agents (IVA2015). Springer, pp 331–334
30. Tutenel T, Bidarra R, Smelik RM, Kraker KJD (2008) The role of semantics in games and simulations. Comput Entertain (CIE) 6(4):57
31. Kessing J, Tutenel T, Bidarra R (2009) Services in game worlds: a semantic approach to improve object interaction. In: Natkin S, Dupire J (eds) 8th international conference on entertainment computing (ICEC 2009), Paris, 3–5 Sept 2009. Springer, Berlin/Heidelberg, pp 276–281
32. Kraayenbrink N, Kessing J, Tutenel T, de Haan G, Bidarra R (2014) Semantic crowds. Entertain Comput 5(4):297–312
33. Ercan S, Harel R, Peperkamp J, Yilmaz U (2014) Virtual humans in games: realistic behavior and emotions for non-player characters. BSc thesis, Intelligent Systems, TU Delft, Delft
34. Hindriks K, van Riemsdijk B, Behrens T, Korstanje R, Kraayenbrink N, Pasman W, de Rijk L (2011) Unreal Goal Bots, In: Dignum F (ed) Agents for games and simulations II. Springer, Berlin, pp 1–18
35. Hudlicka E (2016) Virtual affective agents and therapeutic games. In: Luxton DD (ed) Artificial intelligence in behavioral and mental health care. Elsevier
36. Rizzo A, Shilling R, Forbell E, Scherer S, Gratch J, Morency L-P (2016) Autonomous virtual human agents for healthcare information support and clinical interviewing. In: Luxton DD (ed) Artificial intelligence in behavioral and mental health Care. Elsevier
37. Reisenzein R, Hudlicka E, Dastani M, Gratch J, Hindriks K, Lorini E, Meyer JJC (2013) Computational modeling of emotion: toward improving the inter- and intradisciplinary exchange. Affect Comput IEEE Trans 4(3):246–266

38. Hudlicka E (2014) From habits to standards: towards systematic design of emotion models and affective architectures. In: Bosse T et al. (ed) Emotion modeling: towards pragmatic computational models of affective processes. Springer, pp 3–23
39. Hudlicka E (2003) Modeling effects of behavior moderators on performance: evaluation of the MAMID methodology and architecture. In: Proceedings of 12th conference on behavior representation in modeling and simulation, Phoenix
40. Hudlicka E (2008) Modeling the mechanisms of emotion effects on cognition In: Proceedings of the AAAI fall symposium on biologically inspired cognitive architectures. AAAI Press, pp 82–86

Part III
Applications

Chapter 14
Emotion and Body-Based Games: Overview and Opportunities

Nadia Bianchi-Berthouze and Katherine Isbister

Abstract In this chapter we examine research and theory concerning body movement as a means for expressing emotion, and techniques for recognizing expressions of emotion in the context of games. We discuss body movement as a means for biasing emotional experience and encouraging bonding in social interaction. Finally, we discuss gaps and opportunities for future research. Promising directions include broadening the scope of body-based games and emotion to take proprioception into account, as well as other less explored body channels such as muscle activation and action-related sound.

Introduction

The last two decades have witnessed an increased interest in studying emotion in many research fields (e.g., psychology, neuroscience, computing, engineering, design, medicine, philosophy, HCI). This growing interest in emotion is due to the recent appreciation of the close interaction between cognition and emotion and how emotion is indeed critical to many cognitive processes [1]. Among the various aspects of emotion that are being investigated, of particular interest to game researchers is the relation between emotion and the body as a channel of input and feedback [2]. This interest is driven by the fact that the last 10 years have brought robust body movement-tracking to all of the major game consoles, and have ushered in an era of increasingly sophisticated movement tracking capabilities in smartphones. This has made it possible to create and release games that take advantage of body position and movements, as core mechanics and also as feedback systems for understanding player response. There has been a proliferation of games in this space, and also of research about such games.

N. Bianchi-Berthouze (✉)
UCLIC University College London, Gower Street, London, UK
e-mail: n.berthouze@ucl.ac.uk

K. Isbister
Computational Media, University of California, Santa Cruz, CA, USA
e-mail: katherine.isbister@ucsc.edu

Given the strong relationship between movement and emotion, researchers have been exploring the opportunities that such games might provide. Using terms such as exergames [3] and exertion games [4], researchers have been investigating how physical activity could become more enjoyable and affordable through the use of games in either the comfort of one's home or in the outdoor environment. They have attracted a lot of interest from the research community and from the industry, and are seen as an opportunity to address health challenges (e.g., diabetes) that characterise our society. More broader terms such as body games [5] and movement-based games [6] have been subsequently introduced to consider other benefits of movement such as its general positive effect on emotional and social well-being as well as the opportunities that it offers for the design of the game itself.

For the purposes of this chapter, we use the umbrella term 'body-based games' to take a broad perspective of what the body is, how it expresses and biases our own emotions and the emotions of others, as well as how it can be tracked by sensing technology. Rather than providing a definition of what emotion is and what body-based games are, this chapter aims to provide an overview of the relationships between these two concepts. There is in fact a great deal of research that establishes a link between body positioning and motion, and emotion. And there is a body of game-based research that explores motion as a way to generate emotions in players, in games for entertainment as well as 'serious' purposes. This chapter covers both sets of terrain, building bridges between them and establishing areas for future work.

We begin by examining research and theory concerning body movement as a means for expressing emotion, and techniques for recognizing these expressions in the context of games. Then, we discuss body movement as a means for biasing emotional experience. Next, we cover body movement as a means for bonding in social interaction. Finally, we discuss gaps and opportunities for future research, given this overview. This includes broadening the scope of body-based games and emotion, to take proprioception into account, as well as other less explored body channels such as muscle activation and action-related sound, when considering design and evaluation of games.

The Body as a Means for Expressing Emotions

Theory

Even if the term emotion is commonly used both in everyday life and in research, there is still much debate and disagreement concerning what emotions are [7, 8]. In this chapter, rather than providing an overview of the different definitions of emotion and related affective processes, we focus on how these can be measured or can be regulated. An emotion, or more generally an affective state, is accompanied (or defined) by neural, physiological and behavioural changes triggered in response to the evaluation of an event. For example, research in digitally augmented physical

playgrounds [92] has shown that specific physiological measurements correlate well with and allow quite reliable predictions of the player's preferences while using the playground [93]. They also showed that these characteristics differ to a certain degree from the ones triggered by exertion. Beyond physiological and neural changes, a growing accumulation of research has shown that body expressions (and not just facial expressions) are an important channel for affective communication (see [9] for a review). De Gelder [2] argues that, differently from facial expressions, body expressions tell us not only how a person feels but also how the person is ready to respond (action tendency) to an emotional event. In this chapter, we focus mainly on body movement as an affective modality since it is the one that is used to control the body-based games and that is visible to others.

Two main approaches are used to characterise emotional expressions: dimensional and discrete. Initial work by Ekman and Friesen [10] suggested that the body may be better at communicating broader dimensions of affect than discrete categories. Subsequent studies (e.g., [11–14]) on both acted and naturalistic body expressions confirmed that people do indeed used affective dimensions such as arousal, valence, potency and action tendency when describing these expressions. However, perceptual studies also showed that people do describe body expressions by using discrete emotions with level of agreements well above chance and at levels similar to the ones observed for other modalities (e.g., [15–17]). In fact, a recent study by Aviezer et al. [18] has further confirmed similar results from previous work ([9] for a review) showing that body expressions rather than facial expressions allow us to discriminate between intense positive and negative emotional states.

Various coding models have been proposed to facilitate the analysis of affective body expressions. They aim is to capture the body configuration (its form) as well as its dynamics. Neuroscience studies show that these two types of features may be partially redundant but in combination they help to recognize more complex expressions or solve inconsistencies (e.g., [19–21]). Those studies also show that form alone provides stronger cues for the emotional categorization process than dynamic cues do.

The second main distinction between body description approaches is the use of high-level versus low-level descriptors. The first, often based on the Laban approach [22], describes body expressions through coarse dimensions such as Openness, Jerkiness, Directness. The second approach instead aims at describing body expressions by providing a more precise measure of the distances between body joints and angles between body segments [16, 23]. The review by Kleinmsith et al. [9] provides a set of detailed summary tables of these approaches listing high-level and low-level features and their relationship to both discrete emotions and affective dimensions. While high level descriptors are very useful as they provide a compact description of the expression, the emergence of full-body tracking technology makes low-level descriptions a feasible and possibly richer approach to describe body expressions. A body of work relating these low-level features to emotions is indeed emerging (for an in depth discussion on this topic see also [24]. The study by Kleinsmith et al. [12] show that nuances of emotions can be explained in terms of variation of low-level postural features.

An attempt to provide a more comprehensive multilevel framework for coding is proposed by Fourati et al. [25, 26]. This framework combines both high-level and low-level description of body expressions and includes anatomical, directional and posture/movement descriptors. In this work, they investigate both the use of body tracking technology to provide continuous and rich description of the expressions, as well as qualitative gross descriptors provided by human observers. They argue that such a unified framework is crucial to facilitate the investigation of body expressions of emotion in everyday action and when studying multiple types of action at the same time. Through the use of the framework, they show the existence of a hierarchy of features important in the categorization of emotional expression in everyday action. A feature hierarchy has also been found by Kleinsmith et al. [12, 27] in prototypical body expressions of emotions.

Finally, as with other modalities, factors such as gender, culture and age among others may affect the way we use our body to express an emotion, as well as how we interpret other people's body expressions. A study by Kleinsmith et al. [15] showed that cultural differences exists in the way certain body expressions are interpreted in terms of their valence and arousal level. This is also supported by the more recent work by Volkova et al. [94] on perception of body expressions in story-telling.

Practice

These findings, together with the ready availability of body-based game controllers, have led game designers to consider the opportunities that the body channel offers to personalize the game experience for the player to heighten emotional impact. For example, games that incorporate improvisational movement into the core game mechanic allow for a broad range of emotional expressions. *Yamove*! (Fig. 14.1) is an instance of this approach [28]. Dancers can modulate the emotional tenor of their movements based on the music they are listening to, and how they are feeling (or how they want spectators to feel). *Yamove*! does not track and recognize player emotions—it simply offers a range of expressive possibilities to players. However, other research teams have built on initial work in the field of affective computing, toward building systems able to automatically discriminate between affective body expressions (e.g., early work by Camurri et al. [29] in the dance context; and Bianchi-Berthouze et al. [23] for acted postural expressions). For a more complete review see surveys on automatic perception and recognition of affective expressions [9, 30].

More recently, researchers have started to tackle the problem of recognizing naturalistic body expressions in order to create systems that can be applied to real-life situations. Table 14.1 provides a summary of the studies discussed here, as well as the datasets that were used and that are generally available to the research community upon request from the authors. A study that aims at detecting emotional states from non-acted body expressions in full-body games (Nintendo Wii sport games) is presented in [13]. The aim was to recognize four player's emotional state

Fig. 14.1 *Yamove*! [28] encourages improvisation from players, allowing for a range of emotional expression. The game's core mechanic is improvised dance. Two dance pairs compete against each other in a dance battle. Each pair makes up moves that they can do well together—scoring is based on synchrony of movement, as well as creativity and pace (Image used with permission)

and levels of affective dimensions during replay windows, i.e., when the player is observing and re-evaluating his/her performances in the game. As the context is quite static, the system was built to recognize the affective message conveyed by the configuration of static postures. Full body motion capture sensing technology was used to this purpose. The results showed correct average recognition rate just above 60 % for four affective states (concentrating, defeated, frustrated and triumphant) and 83 % for two affective dimensions (arousal and valence). The results were comparable with human observers' level of agreement reached for the discrete emotions (67 %) over the same set of stimuli and around 85 % for valence and arousal dimensions. In a subsequent study [31], they show that in these semi-static situations, the form features led to performances similar to agreements between human observers even when those were rating the animated clips rather than the apex postures. Moving to a more dynamic situation, Savva et al. [32] investigated the recognition of four emotional states while playing the game. Using dynamic body features, the system performance reached an overall accuracy of 61.1 % comparable to the observers' agreement (61.49 %). Zacharatos et al. [33, 34] repeated similar investigations using different motion capture systems in the context of Microsoft full body Kinect games. In [33], they investigated the possibility of discriminating between low arousal and high arousal states. A vision-based system rather than a mocap system was used to track and measure the body expressions of the player. Laban-informed dynamic features were used to describe the movement. The results from two studies show on average recognition performances just above 90 %. In [34], through postural features captured by the MS Kinect skeleton, they modeled

Table 14.1 Body-movement-based emotion recognition systems in game practice and related naturalistic datasets (over three pages)

ID	Dataset	Emotions	Body tracking	Body features	Target performances	System performances
Kleinsmith et al. [13]	AffectME-posture naturalistic dataset-Nintendo Sports Game	Discrete: concentrated, defeated, frustrated, triumphant / Dimensions: arousal, valence, potency, avoidance	Full body motion capture (Animazoo suit)	All body joint angles at apex of expression / Normalized to feasible movements	Human agreement: 66.7% (discrete emotions), 85% (arousal, valence), poor (potency, avoidance)	Discrete emotions: 63.5%, arousal, valence: 83% / Machine learning: SVM, NLP
Kleinsmith et al. [31]	As above	Discrete: concentrated, defeated, frustrated, triumphant	As above	As above but on 5 frames of a 200 ms windows centred on apex	Human agreement: 66.7% (discrete emotions), 85% (arousal, valence), low (potency, avoidance)	Discrete emotions: 63.5%, arousal, valence: 83% / Machine learning: SVM, NLP
Savva et al. [32]	Affect ME-movement naturalistic dataset: Nintendo Sports Game	High negative, happiness, concentration, low negative	Full body motion capture (Animazoo suit)	Rotation, angular velocity, angular acceleration, direction: hands, head, arms, forearms, spine. Movement amount all 17 joints	Human agreement: 61%	Average: 61.5% / Machine learning: Recurrent Neural Network (RNN)
Gao et al. [90]	Touch-based game	Discrete: excited, relaxed, bored, frustrated. Arousal, valence dimensions	Iphone touch screen	Finger strokes: direction, length, pressure, velocity	Self-reported	Emotions: 77% / Dimensions: 88% / Machine learning: DA, SVM, NLP

Zacharatos et al. [33]	Microsoft kinect game	Meditation, concentration, excitement, frustration	PhaseSpace impulse X2 motion tracking system with 8 cameras	Direction, velocity, acceleration, jerk of feet and hand	Cross-validation	Binary classification: 91% / Four-classes: 85% / Machine learning: NLP
Zacharatos et al. [34]	MS kinect game playing (postures)	Concentrated, defeated/frustrated, triumphant	Kinect sensors	Joint rotations at frame level for all joints of kinect skeleton	Human agreement: 72%	Average: 56.4 / Machine learning: NLP
Olugbade et al. [37, 38]	EmoPain dataset: 3 physical rehabilitation exercises	Healthy people, low pain, high pain	Animazoo full body suit, 4 BTS EMG probe on high and low back	Forms and rotational information: head, arms, lower legs, hips. EMG: general statistics and activation and deactivation points	Self-reported pain level (People with low-back chronic pain and healthy participants)	Stretching forward: 86% / Full trunk flexion: 94%, Sit sit-to-stand: 69% / Machine learning: SVM, RF
Aung et al. [35, 36]	As above but 5 physical exercises and non-instructed movements	Pain-related behaviour: guarding, hesitation, limping, bracing	Same as above	17 body joints measured by motion capture and 4 EMG probes: a large set of form and kinematic features	Human agreement varying according to exercise and pain behaviour: (ICC) 0.5–0.8	Correlations varies according to exercise type: 0.1–0.07 / Machine learning: RF
Griffin et al. [39, 40]	UCL-body laughter: people playing games while standing or sitting (multi-person)	Hilarious laughter, social laughter, awkward laughter, fake laughter, non-laughter	Animazoo suit	All 17 body joints: min, max, range of angular rotation, direction of movement of spine, energy over windows of (non) laughter	Human agreement: 0.94 (Ground truth: ranking)	Laughter types: 0.91 (Ground truth: ranking) / Machine learning: various regression models but best results with RF

(continued)

Table 14.1 (continued)

ID	Dataset	Emotions	Body tracking	Body features	Target performances	System performances
Niewiadomski et al. [42] extended from	Person laughing	Laughter, non-laughter, laughter intensity	Two scenarios: X-sense motion capture, MS kinect	Head speed statistics, trunk leaning and throwing (periodicity, amplitude, impulsiveness), shoulder (energy, periodicity, correlation), etc.	Human observers: 0.71	Average F1: 0.66–0.80
Mancini et al. [41]						Laughter intensity: 0.44 Machine learning: (SVM, RF, SOM, k-NN, probabilistic models)
Fourati et al. [25, 26]	Everyday movement: sitting, walking, knocking, moving objects, lifting, throwing, etc.	Anxiety, pride, joy, sadness, shame, anger, neutral, panic-fear	X-sense motion capture	Multi-level framework kinematic and form features at both low- and high-level	Acted	Performances varies according to emotion and action: 53–92 % Machine learning: RF, SVM

three of the emotional states used in [13] with a performance level of 56.4 % despite the use of a simpler skeleton model. All these studies show performance well above chance levels, and in most cases well close to human agreement as it can be seen in Table 14.1.

Still in the area of exertion body technology, work by Aung et al. [35, 36] and Olugbade et al. [37, 38] investigated the possibility of detecting pain-related behaviour and fear of movement to inform the design of affective-aware technology for gamified physical rehabilitation. By using mainly gross level body features measured from data from mocap suit and EMG sensors worn by the patients, they were able to predict well above chance level the pain level (discretized into none, low, high) self-reported at the end of each physical exercise and pain behaviour (e.g., guarding) as rated by physiotherapists. Though not specifically in a computer game situation, emotions related to playful multi-person interaction are studied in Griffin et al. [39, 40]. They explored the possibility of automatically classifying laughter types from body expressions. Results show that combining form features with energy features led to recognition performances for three laughter types (hilarious, social, fake laughter) and no laughter in both standing and sit-down situations that were comparable with humans' agreement levels. By adding directional form features and kinematic features, the results were further improved. Whilst this work is based on low-level features, Mancini et al. [41] and subsequently Niewiadomski et al. [42] explored high level features of body laughter. An implementation of laughter recognition capabilities in the context of computer games with an artificial co-player is provided in Mancini et al. [43]. The avatar receives multimodal signals from the players to understand when they are laughing and when it is appropriate to laugh and how to laugh (e.g., mimicry) in response with its body [44, 45]. Other multimodal databases including full body movement in a two-person (non-computer) game scenario is reported in [91] aiming to foster research on automatic recognition of social interaction predicates. All these studies provide evidence of a clear increase in focus not only on body as an affective modality but also on the move towards real-life complex situations, a very important step toward being able to deploy such recognition capabilities in real-life applications.

The Body as a Means for Biasing Emotions

Theory

Theories of embodied cognition [46] suggest a dual role of body expressions. Body expressions not only convey to others how we feel, but also affect how we feel and related cognitive processes. As body expressions were recognized to have an important role in communicating emotions, Risking et al. [47] investigated how a person's confidence level could be manipulated by asking them to held a body position that reflected a specific emotional state: a slumped position reflecting

submissiveness and an upright position reflecting confidence. These results were confirmed more recently by the work of Brinol et al. [48] showing that an enacted affective body position biased people's attitudes towards the enacted emotion.

This biasing effect has also been observed in relation to judgment of objects or events a person is asked to evaluate. Early work by Cacioppo et al. [49] observed that arm gestures performed during the evaluation of neutral objects affectively biased their appreciation. Arm gestures that are generally associated to an approach-motivational orientation lead to more positive judgement of the neutral objects than arm gestures associated with an approach-withdrawal orientation. Memory processes are also seen to be facilitated by related affective body expressions. Casasanto and Dijkstra [50] showed that moving objects with upwards facing hands facilitated the retrieval of positive emotions whereas downwards hands led to faster retrieval of negative emotions. Similarly, positively-valenced body movements were shown to lead to be more easily persuaded [51]. Recent work by Carney et al. [52] investigated the biological processes underlying these biasing mechanisms. They found that the production of hormones related to the readiness of an emotional response (e.g., attacking vs. withdrawing) was affected by the enactment of body expressions that reflected such emotional states (highly confident vs highly submissive).

The effect of body expressions on emotion can be also modified by altering the perceptions of one's body. Recent neuroscience studies on sensory feedback integration show that people continuously update the perception of their own body (e.g., [53, 54]). Building on these findings, Tajadura-Jimenez et al. ([55] showed that people's perception of the length of their body and body parts (perceiving longer arms) can be manipulated by altering the sound of one's body action with a consequent effect on people's behaviour and emotional states.

Practice

Building on this body of work, researchers in the field of body-based technology and in particular in game design have started to investigate how such biasing mechanisms could be exploited to design better player experiences. Lindley et al. [56] showed a relationship between 'naturalness' and freedom of movement and the emotional experience of a game. Their study found that an input device that encouraged more 'natural' body movements (i.e., Donkey Konga Bongos) led to an increase in emotional expressions and social interaction. These were measured both in terms of vocal and non-verbal behaviour. Similar results had been previously observed in Berthouze et al. [57] within a different playing context and using a different type of body movement controller. Both studies showed also that the use of body movement control related to the story of the games led players to freely enact other strongly emotionally valenced context-related expressions that could even distract from the main aim of the game, facilitating a broader affective experience.

Pasch et al. [58] and Nijhar et al. [59] build further on these findings showing that the emotional experience a player was looking for led to a different appropriation of the movement recognition precision offered by the game controller. For players motivated to win, the body movements that the game required were used to win the game. Instead, players playing to relax made used of increased recognition precision of the game controller to engage with their own body movement. Melzer et al. [60] and Isbister [61] developed on this further by looking at how body movements may affect the emotional component of the gaming experience. Their studies found that games that encourage body movement leads to higher levels of emotional arousal than those that use a standard controller stand standard controller. Building on this body of work and on the theory of embodied cognition, Berthouze [62] suggests a framework that extends previous engagement models presented in the game literature to include the role of proprioceptive feedback. She proposes five categories of body movements that affect the player's experience: movements necessary to play the game, movements facilitating the control of the game, movements related to the role-play the game offers, affective body expressions and social gestures.

The recent work on sensory integration and body representation update has pushed the boundaries for exergames and their use further into contexts where emotional experience is critical. Singh et al. [63] investigated the use of psychologically-informed movement sonification to change people perception of movement capabilities during chronic musculoskeletal pain physical rehabilitation. When using the proposed tracking and sonifying wearable device, people reported to feel more confident in moving, to perform better (even when this was not the case) and demonstrated higher copying capabilities, i.e., being more ready to take on more difficult challenges [64]. The effect of sound on emotion and behaviour was also demonstrated by Bresin et al. [65]. By altering the sound produced by a person's walking steps, they were able to alter the person's perception of the walking surface material (e.g., snow) and this was reflected in a congruent change in walking style and reported emotional state. In Tajadura-Jimenez et al. [66], the authors showed that through the use of special shoes (Magic Shoes – Fig. 14.2) embedded with microphones to capture and deliver back to people (via headphone) the altered sound of their footsteps, they could control people's perception of their own body e.g., (higher frequency sound made people feel thinner) and alter accordingly their walking behaviour (e.g., faster movement) and their emotional states (more positive).

The Body as a Means for Social Bonding

Theory

Incorporating social interaction into a movement-based game adds a layer of complexity to understanding the emotional impact of the game. Emotional expression
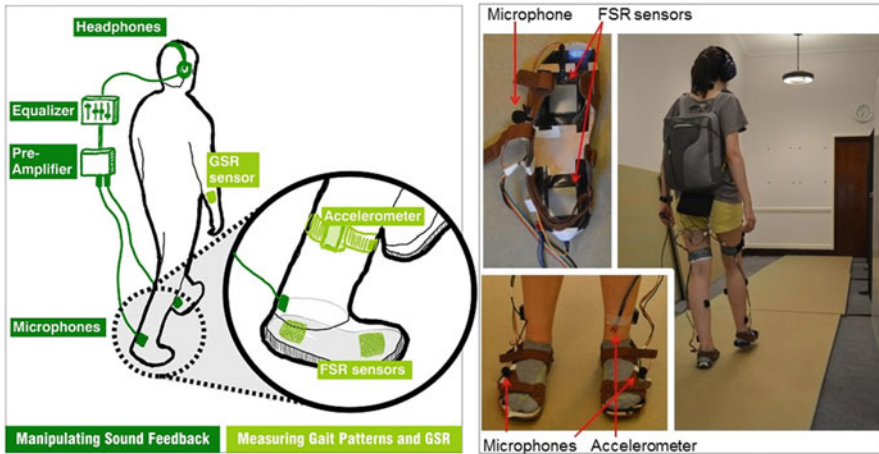
Fig. 14.2 Magic shoes: altering one's body perception through manipulation of sounds made by the sensed body actions [66] with permission of reprint

and signalling is an important aspect of human interaction; therefore in a social play context we must examine emotions as they unfold socially. We need to understand not just the individual's feelings, but also, the effect these have on fellow players and spectators, and vice versa. Researchers have demonstrated that when we observe another person enacting an emotion with the face and/or body, we experience their emotions to some degree—a phenomenon referred to as 'emotional contagion' [67]. Thus, movement mechanics in games that encourage the performance of particular emotional states can be expected to induce some emotional response not just in the player, but also in fellow players and in spectators. Researchers have also found that particular emotional effects can be evoked through encouraging or inducing movement synchrony between people [68, 69]. Specifically, inducing coordinated movement increases compassion and empathy for one another, and social connection to one another. Finally, there is literature that links the manipulation of interpersonal distance—the space between people as they interact—to emotional responses [70, 71]. Bringing people closer together than is socially appropriate in a given cultural context, for example, can lead to strong negative emotions.

Game designers and researchers have developed theory that can be useful in understanding the impact of social movement-based play on emotions, towards designing better social movement and 'exer' games. At a fundamental level, game researchers have postulated that games provide a safe 'magic circle' within which alternate social movement practices are acceptable and even desirable [72, 73]. This can allow interplay between the emotions a person would normally have about an interaction, and how they feel given that what is happening is 'only a game', opening up interesting terrain for exploring and working with emotions that might otherwise be overwhelming or unacceptable. Taking a close look at the social interaction that happens around games, researchers have separated explicitly social play from sociability that is happening in and

around that play [74]. For example, I might give a fellow player a happy 'high five' after winning a round, and that would be sociability, whereas the game Dance Central actually uses a 'high five' performed between opposing players to begin a game-play round—this is 'social play'. In either case, the movement may result in emotions, but the tenor of these emotions could differ depending upon whether the movement was spontaneous or was required in the service of gameplay.

Researchers have also pointed out that not all player movement is accurately detected by many movement-based games, so therefore much of the movement players engage in is actually 'gestural excess' [75] that is not analysed and made use of by the game system. Players often put more movement expressivity than is necessary into movement-based games [62]. As players' emotions can be strongly related to the manner in which they perform the game's movement mechanics [6, 76], cultivating this gestural excess through designing the social framing of the game can be seen as an important component of designing social movement-based games [56, 77]. Researchers have also pointed out that we must consider spectators when we design social games [78]. Games that involve physical performance often generate a spectacle that other potential players observe before playing. So successful design of such games needs to include conscious consideration of the game's emotional effects upon spectators as well as players. Finally, there is work investigating the design values and properties of 'supple' interfaces from the Human Computer Interaction literature, which has relevance to evoking social emotions with movement games [79]. Suppleness is a use quality that is defined as including the use of subtle social signals, emergent dynamics, and a focus on moment-to-moment experience (as opposed to end goals or tasks). Successful movement-based social games may be more likely to have suppleness as a characteristic, and suppleness may be of value in guiding design decisions for movement-based games meant to evoke positive emotions.

Practice

In the past 10 years, there has been a rapid acceleration in the number of movement-based social games created for both research and commercial purposes. This has been facilitated by the introduction of movement-controllers for the major game consoles (Nintendo Wii, Sony Move, Microsoft Xbox Kinect), and by the rapid spread of sensor-enabled smart phones and increased bandwidth of network connections (see for example the indie game *Bounden*, Fig. 14.3). Researchers interested in the emotional effects of social movement games have been able to use these platforms and other readily available components and hardware elements to construct games with which to study the impact of social movement mechanics. Some examples include The Exertion Games Lab's *I-dentity* and *Musical Embrace* [80, 81]; the NYU Game Innovation Lab's *Wriggle*, *Yamove*!, and *Pixel Motion* [61, 82, 83]; and the *Oriboo* [5]; and the socially aware interactive playground work done at Twente University [84].

Fig. 14.3 *Bounden* (2015) is a smartphone-based game that requires two players to each have their thumb on the screen, working together to keep a virtual sphere visible and move it through a path of rings by tilting and rotating the device together. The moves that result were actually choreographed by the Dutch National Ballet, ensuring a (somewhat) graceful result; image used with permission

There has also been some effort to aggregate findings about the design and impact of social movement games. Mueller and Isbister engaged in an aggregation of best design practices in the form of ten movement game guidelines [6, 76], which include information about designing to facilitate social fun. Márquez-Segura and Isbister wrote a chapter aggregating recent research on co-located physical social play, which includes detailed descriptions of the *Yamove*! the *Oriboo* systems and accompanying research work [77]. This chapter highlights the importance of making the best use of technology, setting, and players as design material; allowing for and embracing player influence and impact when shaping gameplay; and encouraging and protecting the 'we' in social play.

Future Work

Improving Sensing of Emotional Cues

As discussed in section "The Body as a Means for Expressing Emotions", systems are becoming capable of interpreting the emotional content expressed through non-verbal behavior, including body expressions. However, this capability has not yet been extensively engaged by computer games, though examples have begun to appear. As sensors become cheaper and ubiquitous, there is still the need to fully understand which affective dimensions can be captured through the set of sensors

available (e.g., full body motion capture system vs smartphone), especially when these provide minimal datapoints.

At the same time, as games are ubiquitous, it is also time to consider that the sensing technology available to track people's body expressions may not be predefined, and that different devices may be available at different stages of a game's life cycle. This is particularly important not just in the entertainment context. A recent study by Singh et al. [63] show that gamified physical rehabilitation should be designed with a mobile and ubiquitous model in mind, to facilitate transferring of skills from physical exercise sessions to everyday functional movements. In addition, the social aspects also suggest and invite researchers to consider measuring not just individual players' emotions but group emotions as well as audience emotions. For example, new body-based measures able to capture the level of bonding within the group, and congruency of emotions between people in the group are needed. It could be interesting to detect the emotion group leader and support that person in altering or regulating the emotional states of the group.

Most work is still focused on measuring visible body expression, missing important information that is not easy to track using motion sensors. For example, tension of the muscles in the arms, which may indicate readiness to act in a specific way. Work by Huis in t'Veld [85, 86] has shown the existence of muscle activity patterns that relate to particular emotions. This work is still very preliminary and calls for using games as ways to study this relationship and to exploit it toward better personalization of the game experience. Finally, an affective channel not fully exploited in the game context even if ubiquitously present, and increasingly finely measured, is touch behaviour. A large body of research shows that touch is a powerful affective modality through which people express their emotion, communicate emotions to others [87] and also express what they feel about objects [88]. For example, Sae-Bae et al. [89] showed that touch-based authentication gestures were more pleasurable as well as more secure than standard text-based passwords. Gao et al. [90] shows that using touch behaviour during a touch-based smartphone game, the system could detect people's affective states with very high performance (see Table 14.1 for details).

Improving Body-Based Game Evaluation for Social Games

In studying movement-based social games, researchers use game logs, video recordings, interviews, and post-play surveys to understand impacts on players. Understanding of players' emotions has been a small part of the overall set of research questions and measures in extant studies, and so there is as yet very little detailed information about how to evoke particular social emotions given particular design choices. Designers need to be able to unpack at a reasonably granular level what is happening emotionally and when for players, so they can build emotional evaluation into prototyping and iteration of these games. So far, this

is time consuming and difficult—it can take many hours to code video logs, and not all emotions are legible using these records. Self-report of emotion during gameplay disrupts the experience and post-surveys and interviews can only give more fuzzy, aggregate impressions [77]. There is a continuing need for more sophisticated evaluation techniques for capturing the nuances of social emotions during gameplay. Ideally, researchers could use some combination of unobtrusive physiological and self-report measures, triangulated with game log data, to get a good picture of what is happening emotionally for players and why.

In terms of design practices, there is a continuing need for dialog between commercial game developers and academic researchers, toward capturing craft-based tacit knowledge and propagating it more broadly to future social movement game designers, including those in the games for impact sector who must make their design process and criteria more explicit, tethering these choices to desired outcomes [6, 76].

Conclusions

This chapter provided an overview of the current state of the art in understanding body-based emotion cues and their use in exergames and other body-based game design. Much research has been conducted in support of reading basic emotional signals from the body, and some progress has been made in incorporating this knowledge into game design choices. There are substantial future opportunities for broadening emotion sensing capabilities, designing emotion into body-based games, and evaluating their impact. To date, there has not been much crossover between research communities considering input sensors and game outputs, and those that consider affective user experience sensing and analysis. In future, it could be fruitful to merge these lines of thought, in order to more richly understand what is happening for players, and become more methodical and sophisticated about designing body-based effects in games, including therapeutic and other 'serious' games uses (for example, sensory integration, pain management, and self perceptions related to the body). Merging these perspectives might also allow for increased sophistication in evoking complex emotions, enabling the examination of and design for higher order constructs such as creativity and team feeling.

References

1. Damasio AR (1994) Descartes' error: emotion, reason and the human brain. Avon Books, New York
2. de Gelder B (2009) Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. Philos Trans R Soc 364(3):3475–3484
3. Sinclair J, Hingston P, Masek M (2007) Considerations for the design of exergames. Proc GRAPHITE'07 1:289–295

4. Mueller FF, Edge D, Vetere F, Gibbs MR, Agamanolis S, Bongers B, Sheridan JG (2011) Designing sports: a framework for exertion games. In: Proceedings of CHI 2011, pp 2651–2660

5. Márquez Segura E, Waern A, Moen J, Johansson C (2013) The design space of body games: technological, physical, and social design. In: Proceedings of CHI 2013, Paris, France

6. Isbister K, Mueller F (2015) Guidelines for the design of movement-based games and their relevance to HCI. Hum Comput Interact, Spec Issue Games HCI 30(3–4):366–399

7. Kleinginna PR, Kleinginna AM (1981) A categorized list of emotion definitions with suggestions for a consensual definition. Motiv Emot 5:345–379

8. Scherer KR (2005) What are emotions? And how can they be measured? Soc Sci Inf 44(4): 693–727

9. Kleinsmith A, Bianchi-Berthouze N (2013) Affective body expression perception and recognition: a survey. IEEE Trans Affect Comput 4(1):15–33

10. Ekman P, Friesen W (1967) Head and body cues in the judgment of emotion: a reformulation. Percept Mot Skills 24:711–724

11. Paterson HM, Pollick FE, Sanford AJ (2001) The role of velocity in affect discrimination. In: Proceedings of 23rd annual conference of the cognitive science society, Lawrence Erlbaum Associates, pp 756–761

12. Kleinsmith A, Bianchi-Berthouze N (2005) Grounding affective dimensions into posture features. LNCS: Proceedings of 1st international conference on on affective computing and intelligent interaction, pp 263–270

13. Kleinsmith A, Bianchi-Berthouze N, Steed A (2011) Automatic recognition of non-acted affective postures. IEEE Trans Syst Man Cybern Part B 41(4):1027–1038

14. Karg M, Kuhnlenz K, Buss M (2010) Recognition of affect based on gait patterns. IEEE Trans Syst Man Cybern Part B 40(4):1050–1061

15. Kleinsmith A, De Silva R, Bianchi-Berthouze N (2006) Cross-cultural differences in recognizing affect from body posture. Interact Comput 18(6):1371–1389

16. Coulson M (2004) Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. J Nonverbal Behav 28:117–139

17. Wallbott HG (1998) Bodily expression of emotion. Eur J Soc Psychol 28:879–896

18. Aviezer H, Trope Y, Todorov A (2012) Body cues, not facial expressions, discriminate between intense positive and negative emotions. Science 338:1225–1229, Issue 30

19. Lange J, Lappe M (2007) The role of spatial and temporal information in biological motion perception. Adv Cogn Psychol 3(4):419–428

20. Atkinson AP, Dittrich WH, Gemmell AJ, Young AW (2007) Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. Cognition 104(1):59–72

21. Roether C, Omlor L, Christensen A, Giese MA (2009) Critical features for the perception of emotion from gait. J Vis 8(6):1–32, No 15

22. Laban R (1980) The mastery of movement. Plymouth, UK

23. Bianchi-Berthouze N, Kleinsmith A (2003) A categorical approach to affective gesture recognition. Connect Sci 15(4):259–269

24. Dael N, Bianchi-Berthouze N, Kleinsmith N, Mohr C (2015) Measuring body movement: current and future directions in proxemics and kinesics. In: The APA handbook of nonverbal communication. APA

25. Fourati N, Pelachaud C (2015) Multi-level classification of emotional body expression. In: International conference on automatic face and gesture recognition, pp 1–8. 10.1109/FG.2015.7163145

26. Fourati N, Pelachaud C (2015) Relevant body cues for the classification of emotional body expression in daily actions. In: IEEE proceedings of international conference on affective computing & intelligent interaction

27. Kleinsmith A, Bianchi-Berthouze N (2007) Recognizing affective dimensions from body posture, vol 4738, Lecture Notes in Computer Science. LNCS. Springer, Berlin, pp 48–58

28. Isbister K (2012) How to stop being a buzzkill: designing Yamove!, a mobile tech mash-up to truly augment social play. Proceedings of MobileHCI

29. Camurri A, Mazzarino B, Volpe G (2004) Analysis of expressive gesture: the EyesWeb expressive gesture processing library. Gesture-Based Commun Hum-Comput Interact LNCS 2915:460–467

30. Zacharatos H, Gatzoulis C, Chrysanthou Y, Aristidou A (2014) Automatic emotion recognition based on body movement analysis: a survey. IEEE Comput Graph Appl 34(6):35–45

31. Kleinsmith A, Bianchi-Berthouze N (2011) Form as a cue in the automatic recognition of non-acted affective body expressions, vol 6975, LNCS. Springer, Memphis, pp 155–164

32. Savva N, Scarinzi A, Bianchi-Berthouze N (2012) Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience. IEEE Transact Comput Intell AI Games 4(3):199–212

33. Zacharatos H, Gatzoulis C, Chrysanthou Y, Aristidou A (2013a) Emotion recognition for exergames using Laban Movement analysis. In Proceedings of Motion on Games (MIG '13). ACM, New York, Article 39

34. Zacharatos H, Gatzoulis C, Chrysanthou G (2013b) Affect recognition during active game playing based on posture skeleton data, In: 8th international conference on computer graphics theory and applications

35. Aung MS, Bianchi-Berthouze N, Watson P, C de C Williams A (2014) Automatic recognition of fear-avoidance behavior in chronic pain physical rehabilitation. PervasiveHealth '14 Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare, pp 158–161

36. Aung MSH, Kaltwang S, Romera-Paredes B, Martinez B, Singh A, Cella M, Valstar M, Meng H, Kemp A, Shafizadeh M, Elkins AC, Kanakam N, de Rothschild A, Tyler N, Watson PJ, Williams AC de C, Pantic M, Bianchi-Berthouze N (n.d.) The automatic detection of chronic pain-related expression: requirements, challenges and a multimodal dataset in IEEE Trans. Affective Computing (in press)

37. Olugbade T, Aung MSH, C de C Williams A, Bianchi-Berthouze N (2014) Bi-modal detection of painful reaching for chronic pain rehabilitation systems. ICMI'14, pp 455–458

38. Olugbade TA, Bianchi-Berthouze N, Marquardt N, Williams A CdeC (2015) Pain level recognition using kinematics and muscle activity for physical rehabilitation in chronic pain. IEEE proceedings of international conference on affective computing & intelligent interaction, pp 243–249

39. Griffin HJ, Aung MSH, Romera-Paredes B, McKeown G, Curran W, McLoughlin C, Bianchi-Berthouze N (2013) Laughter type recognition from whole body motion. In: IEEE proceedings of international conference on affective computing & intelligent interaction, pp 349–355

40. Griffin HJ, Aung MSH, Romera-Paredes B, McLoughlin C, McKeown G, Curran W, Bianchi-Berthouze N (2015) Perception and automatic recognition of laughter from whole-body motion: continuous and categorical perspectives. IEEE Transact Affect Comput 6:165–178

41. Mancini M, Varni G, Niewiadomski R, Volpe G, Camurri A (2014) How is your laugh today? In: CHI'14 extended abstracts on human factors in computing systems, pp 1855–1860

42. Niewiadomski R, Mancini M, Varni G, Volpe G, Camurri A (n.d.) Automated laughter detection from full-body movements. IEEE Transactions on Human-Machine Systems (in press)

43. Mancini M, Ach L, Bantegnie E, Baur T, Bianchi-Berthouze N, Datta D, Ding Y, Dupont S, Griffin HJ, Lingenfelser F, Niewiadomski R, Pelachaud C, Pietquin O, Piot B, Urbain J, Volpe G, Wagner J (2014) Laugh when you're winning, innovative and creative developments in multimodal interaction systems. Springer, Berlin, pp 50–79

44. Niewiadomski R, Mancini M, Ding Y, Pelachaud C, Volpe G (2014) Rhythmic body movements of laughter. In: Proceedings of the 16th international conference on multimodal interaction, pp 299–306

45. Griffin H, Varni G, Tome-Lourido G, Mancini M, Volpe G, Bianchi-Berthouze N (2015) Gesture mimicry in expression of laughter. ACII'15

46. Niedenthal PM (2007) Embodying emotion. Science 316(5827):1002–1005
47. Riskind JH, Gotay CC (1982) Physical posture: could it have regulatory or feedback effects on motivation and emotion? Motiv Emot 6(3):273–298
48. Brinol P, Petty R, Wagner B (2009) Body posture effect on self-evaluation: a self-validated approach. Eur J Soc Psychol 39:1053–1064
49. Cacioppo JT, Priester JR, Berntson GG (1993) Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. J Pers Soc Psychol 65: 5–17
50. Casasanto D, Dijkstra K (2010) Motor action and emotional memory. Cognition 115(1): 179–185
51. Wells GL, Petty RE (1980) The effects of overt head movements on persuasion: compatibility and incompatibility of responses. Basic Appl Soc Psychol 1:219–230
52. Carney DR, Cuddy AJC, Yap AJ (2010) Power posing: brief nonverbal displays affect neuroendocrine levels and risk tolerance. Psychol Sci 21(10):1363–1368
53. Botvinick M, Cohen J (1998) Rubber hands 'feel' touch that eyes see. Nature 391:756, Issue *39*
54. Tsakiris M (2010) My body in the brain: a neurocognitive model of body ownership. Neuropsychologia 48(3):703–712
55. Tajadura-Jimenez A, Tsakiris M, Marquardt T, Bianchi-Berthouze N (2015) Action sounds update the mental representation of arm dimension: contributions of kinaesthesia and agency. Front Psychol 6:689
56. Lindley SE, Le Couteur J, Bianchi-Berthouze N (2008) Stirring up experience through movement in game play: effects on engagement and social behaviour. In: SIGCHI conference on human factors in computing systems, pp 511–514
57. Berthouze NK, Kim W, Darshak P (2007) Does body movement engage you more in digital game play? And why? Lect Notes Comput Sci LNCS 4738:102–113
58. Pasch M, Bianchi-Berthouze N, van Dijk B, Nijholt A (2009) Movement-based sports video games: investigating motivation and gaming experience. Entertain Comput 1(2):49–61
59. Nijhar J, Bianchi-Berthouze N, Boguslawski G (2012) Does movement recognition precision affect the player experience in exertion games? Int Conf Intell Technol Interact Entertain (INTETAIN) LNICST 78:73–82
60. Melzer A, Derks I, Heydekorn J, Steffgen G (2010) Click or strike: realistic versus standard game controls in violent video games and their effects on agression. In: Yang HS, Malaka R, Hoshino J, Han JH (eds) 9th International Conference, ICEC 2010. Springer, Berlin
61. Isbister K, Schwekendiek U, Frye J (2011a) Wriggle: an exploration of emotional and social effects of movement. In: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, pp 1885–1890
62. Bianchi-Berthouze N (2013) Understanding the role of body movement in player engagement. Hum Comput Interact 28(1):42–75
63. Singh A, Piana S, Pollarolo D, Volpe G, Varni G, Tajadura- Jiménez A, Williams A CdeC, Camurri A, Bianchi-Berthouze N (n.d.) Go-with-the-flow: tracking, analysis and sonification of movement and breathing to build confidence in activity despite chronic pain. HCI (in press)
64. Singh A, Klapper A, Jia J, Fidalgo A, Tajadura-Jimenez A, Kanakam N, Bianchi-Berthouze N, Williams A (2014) Motivating people with chronic pain to do physical activity: opportunities for technology design. In: Proceedings of the 32nd ACM conference on human factors in computing systems, pp 803–2012
65. Bresin R, de Witt A, Papetti S, Civolani M, Fontana F (2010) Expressive sonification of footstep sounds. Proceedings of ISon
66. Tajadura Jimenez A, Basia M, Deroy O, Fairhurst M, Marquardt N, Bianchi-Berthouze N (2015) As light as your footsteps: altering walking sounds to change perceived body weight, emotional state and gait. In: Proceedings of the SIGCHI conference on human factors in computing system

67. Hatfield E, Cacioppo JT, Rapson RL (1994) Emotional contagion. Cambridge, Cambridge University Press
68. Marsh KL, Richardson MJ, Schmidt RC (2009) Social connection through joint action and interpersonal coordination. Top Cogn Sci 1(2):320
69. Valdesolo P, DeSteno D (2011) Synchrony and the social tuning of compassion. Emotion 11(2):262
70. Hall ET (1968) Proximics. Curr Anthropol 9(2–3):83
71. Knapp ML, Hall JA (2002) Nonverbal communication in human interaction, 3rd edn. Holt, Rinehart & Winston, New York
72. Huizinga J (1955) Homo Ludens: a study of the play-element in culture. Beacon Press, Boston
73. Salen K, Zimmerman E (2004) Rules of play: game design fundamentals. MIT Press, Massachusetts
74. Stenros J, Paavilainen J, Mäyrä F (2009) The many faces of sociability and social play in games. In: Proceedings of the 13th International MindTrek Conference: everyday life in the Ubiquitous Era, MindTrek'09. ACM, New York, pp 82–89
75. Simon B (2009) Wii are out of control: bodies, game screens and the production of gestural excess. Loading 3(4). http://journals.sfu.ca/loading/index.php/loading/article/viewArticle/65
76. Mueller F, Isbister K (2014) Movement-based game guidelines. In: Proceedings of the SIGCHI conference on human factors in computing systems
77. Márquez Segura E, Isbister K (2015) Enabling co-located physical social play: a framework for design and evaluation. In: Bernhaupt R (ed) Game user experience evaluation, Springer, Switzerland, pp 209–238. http://www.springer.com/us/book/9783319159843
78. Reeves S (2011) Designing interfaces in public settings: understanding the role of the spectator. Human-computer interaction. Springer, Berlin
79. Isbister K, Höök K (2009) On being supple: in search of rigor without rigidity in meeting new design and evaluation challenges for HCI practitioners. In: Proceedings of the SIGCHI conference on human factors in computing systems
80. Huggard A, De Mel A, Garner J, Toprak C, Chatham A, Mueller F (2013) Musical embrace: understanding a socially awkward digital play journey. DiGRA, Atlanta, Georgia
81. Garner J, Wood G, Pijnappel S, Murer M, Mueller F (2014) I-dentity: innominate movement representation as engaging game element. In: Proceedings of the SIGCHI conference on human factors in computing systems
82. Isbister K (2012) How to stop being a buzzkill: designing Yamove!, a mobile tech mash-up to truly augment social play. MobileHCI'12
83. Robbins H, Isbister K (2014) Pixel motion: a surveillance camera enabled public digital game. In: Proceedings of foundations of digital game
84. Moreno A, van Delden R, Poppe R, Reidsma D (2013) Socially aware interactive playgrounds. Pervasive Comput 12:40–47
85. Huis in t'Veld EMJ, Van Boxtel GJM, de Gelder B (2014) The body action coding system I: muscle activations during perception and activation of emotion. Soc Neurosci 9(3):249–264
86. Huis in t'Veld EMJ, Van Boxtel GJM, de Gelder B (2014) The body action coding system II: muscle activations during perception and activation of emotion. Behav Neurosci 8:330
87. Hertenstein MJ, Holmes R, Mccullough M, Keltner D (2009) The communication of emotion via touch. Emotion 9(4):566–573
88. Atkinson D, Orzechowski P, Petreca B, Bianchi-Berthouze N, Watkins P, Baurley S, Padilla S, Chantler M (2013) Tactile perceptions of digital textiles: a design research approach. In: Proceedings of the SIGCHI conference on human factors in computing systems
89. Sae-Bae N, Ahmed K, Isbister K, Memon N (2012) Biometric-rich gestures: a novel approach to authentication on multi-touch devices. In:Proceedings of the SIGCHI conference on human factors in computing systems
90. Gao Y, Bianchi-Berthouze N, Meng H (2012) What does touch tell us about emotions in touchscreen-based gameplay? ACM Transact Comput Hum Interact 19(14):31
91. Siddiquie B, Amer M, Tamrakar A, Salter D, Lande B, Mehri D, Divakaran A (2015) Tower game dataset: a multimodal dataset for analyzing social interaction predicates. ACII'15

92. Lund HH, Klitbo T, Jessen C (2005) Playware technology for physically activating play. Artif Life Robot J 9:165–174
93. Yannakakis GN, Hallam J (2008) Entertainment modeling through physiology in physical play. Int J Hum-Comput Stud 66(10):741–755
94. Volkova E (2014) Expressions in narrative scenarios and across cultures. PhD thesis, Tubingen University. https://publikationen.uni-tuebingen.de/xmlui/handle/10900/58044

Chapter 15
Games for Treating and Diagnosing Post Traumatic Stress Disorder

Christoffer Holmgård and Karen-Inge Karstoft

Abstract This chapter describes the use of games for addressing Post Traumatic Stress Disorder, a syndrome with a strong emotional component, characterized by among other things hyper-arousal. A number of games and game-like tools for treating Post Traumatic Stress Disorder are described. Subsequently, the chapter describes the design, development, and testing of a specific game for addressing Post Traumatic Stress Disorder which uses emotion recognition to characterize and target patient treatment. In clinical testing the game is found to elicit stress in patients suffering from Post Traumatic Stress Disorder based both on self-reports and physiological indicators of stress and arousal. Further, it is shown that features extracted from the physiological signals can be combined with patient background information to predict measures of Post Traumatic Stress Disorder severity. This suggests that the combination of games and measurement of physiological indicators of emotional responses holds potential for creating novel tools for treating and diagnosing mental health disorders.

Introduction

In this chapter, we describe the use of games and game-like artefacts to address the condition known as Post Traumatic Stress Disorder. We describe a study aimed at measuring and changing emotional responses through a custom developed game: *StartleMart*. The *StartleMart* game is a virtual environment tailored for producing emotionally salient experiences for patients suffering from Post Traumatic Stress Disorder. The game exposes patients, veterans traumatized during service deployment, to stimuli carefully crafted in collaboration with domain experts: veterans, psychologists, and psychiatrists. Based on detailed recordings of in-game parameters the game uses physiological indicators of stress to identify

C. Holmgård (✉)
Department of Computer Science and Engineering, New York University, New York, NY, USA
e-mail: christoffer@holmgard.org

K.-I. Karstoft
Forsvarets Veterancenter, Ringsted, Denmark
e-mail: kareningekarstoft@gmail.com

which stimuli have the largest emotional impact on each individual veteran and to estimate syndrome severity. These *personal stress response profiles* then allow for future selection in stimuli, allowing for an individualized presentation of the most efficacious stimuli for each veteran.

In the following sections, we first describe the syndrome of Post Traumatic Stress Disorder (PTSD). We move on to provide an overview of a number of game-based approaches to the treatment of PTSD. Then, we describe how our own game environment, *StartleMart*, was designed to address PTSD, and how emotion recognition was designed as an integral part of the game system. From there, we describe the special characteristics of physiological manifestations of emotional responses in PTSD sufferers, how we used these characteristics to tailor the emotion-recognizing features of StartleMart, and the experiments we designed to test these capabilities. Finally, we present the results from our study and discuss their implications for the future design and use of games for addressing mental health through emotion recognition and stimulation.

Posttraumatic Stress Disorder

Posttraumatic Stress Disorder (PTSD) is a debilitating mental health disorder that occurs in some individuals after exposure to a traumatic event [16]. Exposure to combat places military personnel at significant risk for PTSD [10], resulting in a significant minority of deployed military personnel developing severe PTSD psychopathology that can be unremitting across the life course [27] and cause severe disability.

Hallmark symptoms of PTSD include re-experiencing the traumatic event (i.e. nightmares, intruding thoughts, flashbacks and physiological arousal when reminded of the event), fear-based avoidance of potential triggers (i.e. making an effort to avoid thoughts and external reminders of the event), emotional numbing, and physiological hyperarousal (i.e. being easily startled, being tense or on edge, and sleep and concentration problems). The combination of re-experiencing, avoidance, and hyperarousal symptoms have been found to be efficiently targeted in exposure-based cognitive therapy [9].

The need for effective treatment for PTSD is evident, however, a large proportion of individuals with PTSD do not benefit from the gold standard psychotherapeutic treatment. Therefore, new approaches to treatment of PTSD should be developed, tested and validated [7].

Games for Mental Health and PTSD

Recent years has seen increasing interest in and research on applying computer games for health related purposes, including mental health [15]. Games

and game-like artefacts have successfully been used as mental health interventions by appropriating commercial games [11] and by developing specialized solutions [13].

Specifically for PTSD, efforts in applying games have been centered around two approaches: psychoeducation with promotion of help-seeking behavior and variations of exposure therapy. In the case of psychoeducation, games such as *Family of Heroes* have been shown to increase the occurrence of help seeking in American veterans experiencing post-deployment stress [1]. The game educates veterans as well as family members about post-deployment stress and appropriate responses, helping the veteran take initiative to seek help when necessary and/or helping family members support this process. For exposure therapy, most work has been centered on using computer game engines and virtual reality technology to facilitate exposure therapy.

Virtual Reality Therapy for Treating PTSD

In recent years, Virtual Reality (VR) perspectives have found their way into prevention and treatment of PTSD following military combat. Earlier research has demonstrated the usefulness of virtual environments for treating veterans' PTSD with *virtual reality therapy*, an extension of exposure therapy [12, 19, 33].

More specifically, VR has been suggested as a potent mean for conducting exposure-based therapy [25], as well as for the training and resilience building of soldiers before deployment through Stress Inoculation Training [24, 31]. Notable examples are the *Virtual Iraq* and *Virtual Afghanistan* applications that have shown promising results in clinical testing [22, 23].

It is known from the PTSD literature that exposure to standardized audiovisual trauma cues causes heightened physiological arousal seen as higher heart rate (HR) and elevated Skin Conductance Response (SCR) in PTSD patients compared to controls [21]. VR provides an excellent means for mimicking and cueing the original trauma, thereby creating physiological arousal assumed to be somewhat similar to that created by exposure to the actual trauma stimuli. This VR-induced physiological arousal can then be targeted through well-validated cognitive techniques by a trained clinician [19].

One problematic issue in PTSD is the generalized stress response that in the aftermath of deployment occurs not only in response war-related triggers, but also in everyday situations that may not be similar to the original war trauma. While soldiers after home coming from deployment are most often not confronted with war-like stimuli in their everyday life, they continue to experience heightened arousal when performing everyday activities [17]. Therefore, targeting arousal arising in everyday situations might be even more important than targeting stress responses arising in the context of war-related triggers.

Here, we aim to target the stress-response arising in relation to an everyday activity combined with stress-response arising to specific war-situations from a deployment zone. More specifically, we investigate the physiological arousal arising when combat soldiers with a PTSD-diagnosis play a computer game centered on grocery shopping in a supermarket, and immersed with flashback scenes from Afghanistan. For hyperarousal assessment, we record skin conductance response (SCR) throughout the gaming session and in relation to specific events in the game.

StartleMart: A Virtual Scenario for Assessing Stress Response in PTSD

Shopping for groceries is a daily task reported to be challenging for individuals suffering from PTSD [14]. Supermarkets are rich, unpredictable environments with multiple visual, auditory, olfactory and tactile stimuli. Further, avoiding social and physical contact with other grocery shoppers is virtually impossible. This multifaceted impact from the environment is perceived as stressful for many individuals with PTSD. Hence, to create a game environment that is likely to induce everyday stress reactions, we build the game scenario as a virtual supermarket (see Fig. 15.1a, c, and e).

To target the hallmark PTSD symptoms of re-experiencing, fear-avoidance, and hyperarousal, the game includes a number of stressful events. Re-experiencing of the traumatic event is targeted through flashback-like scenarios, i.e. with the aim of giving the player a feeling of being back in the traumatic situation. More specifically, we present three different war scenarios expected to mimic typical distressing experiences from war zones (see Fig. 15.1b, d, and f). Fear-avoidance behavior is targeted in the game by designing the supermarket with hidden angles preventing the player from gaining full overview of the location. Co-shoppers wander around and will sometimes block the way of the player. One co-shopper acts angrily towards a child, other co-shoppers are staring angrily at the player, and one is walking aggressively towards the player. To address hyperarousal, a dog is barking aggressively at the entrance to the supermarket, and during the game, at one or more occasions, a loud sound of breaking glass is heard.

To support player immersion, we adopt a first-person perspective, with the player acting as a shopper in the supermarket. Specifically, the player walks around the supermarket and is instructed via a shopping list to pick up and eventually buy a number of items. This ensures that the player is moves around and visits most locations in the virtual supermarket. At set locations in the supermarket, the flashback scenarios are activated. Further, a countdown watch is visible to the player at all times in the upper right corner of the screen.

Fig. 15.1 The three traumatic experience cues of the game (**b**, **d**, **f**) and the immediately preceding stressful scenes from everyday life (**a**, **c**, **e**). Elements of the everyday life scenes bleed into the cue scenes, referencing re-experience, a symptom typical for PTSD. (**a**) Sound of ventilator blowing overhead. (**b**) Sound of wind blowing. (**c**) Man walking angrily toward player. (**d**) Afghan running toward player. (**e**) Man staring at player. (**f**) Wounded soldier staring at player

Assessment of Physiological Arousal, Subjective Stress Response, and PTSD Symptomatology

To gain thorough insight on the psychological as well as physiological arousal experienced by the player during the game session, we include subjectively reported distress as well as physiologically recorded arousal. Before starting the game, the player is asked to anchor a scale from 0 to 100 by considering an event

or situation that evokes "no stress or anxiety, completely relaxed" (0), an event or situation that evokes "the highest level of stress or anxiety you have ever experienced, with significant physiological symptoms" (100), and an event that evokes a state in between the two, i.e. "moderate stress or anxiety" (50).

Immediately before the session, the player is asked to rate on this personally anchored scale how distressed he is feeling. Immediately after each round, the player is asked to report how high on the scale his stress level was when it peaked during the game, and what event elicited that response. Finally, the player is asked to report the current level of stress immediately after finishing the round.

To assess physiological arousal, we measure skin conductance (SC). SC is a useful measure of stress and has previously been used to asses soldier stress [20]. Activation of sweat glands is due to activation of the sympathetic nervous system and linked to the reaction to threats. Hence, SC is related to emotional states such as fear and anxiety, or, in more general terms, arousal.

We expect to see relations between PTSD severity and measures of subjective distress and physiological arousal during the game. Specifically, we expect a positive correlation between level of PTSD symptomatology, subjective distress, and physiological arousal. To assess PTSD symptom severity, all participants fill out a self-report measure of PTSD, the PTSD Checklist prior to playing the game. The PTSD checklist is a questionnaire with 17 items corresponding to the diagnostic criteria for PTSD in DSM-IV [2, 30].

In summary, we use SC to obtain information on sympathetic nervous system activation. Further, we use the subjectively obtained information to evaluate agreement between subjectively perceived distress and objectively obtained physiological arousal.

Clinical Trials Using StartleMart

In order to evaluate StartleMart's ability to support the assessment of stress responses a clinical trial was conducted. The trial had the objective of testing whether physiological indicators of arousal may be used to predict PTSD symptom severity as measured by PCL.

Participants and Inclusion Criteria

Since the game was designed specifically to elicit stress in veterans suffering from PTSD, and the literature shows that their stress response patterns should be expected to be significantly different from veterans not suffering from PTSD, a clinical trial including actual PTSD patients was necessary. Thirteen male PTSD patients,

veterans from Danish Armed Forces' military operations in Afghanistan were recruited for the clinical trial. All patients qualified for the PTSD diagnosis, were undergoing regular psychiatric treatment, and were only admitted into the trial after careful evaluation by their regular psychiatrist. All subjects were medicated with Selective Serotonin Reuptake Inhibitors (SSRI) which are known to generally lower sympathetic activity [28]. This means that any recorded physiological responses are assumed to be equal to or lower than the expected responses from unmedicated patients.

Collected Features

A number or features were collected from the participants, falling in five different categories: patient profile features, session features, event behavioral features, self-reported stress response features, and SC stress response features collected from the experimental sessions. The complete feature set for each patient, session, and event is outlined in Table 15.1 along with an indication of each feature's data type and values.

Patient Profile Features

For building the patient profile the patient was subjected to the PTSD Module of the Structured Clinical Interview for the DSM (SCID) [8] and completed the military version of the PTSD Checklist-IV (PCL-M) [5], a 17-item questionnaire that yields a PTSD symptom severity score in the interval 17–85. All patients were also profiled in terms of *age*, number of deployments (i.e. war missions) experienced N_{dep}, and the number of days since their return from their latest deployment N_{day}. The average, standard deviation and range values of the PTSD profile features across all patients are presented in Table 15.2. For the veteran PTSD patients, traumatized by experiences during deployment in this study, we assume that N_{day} may be considered an adequately precise measure of the time passed since the traumatizing experience. The deployment situation as a whole may be considered a highly stressful experience and as such part of the traumatizing situation. This means that the age of the trauma for all purposes here is assumed to be equivalent to N_{day}.

Session Features

Session features simply indicate what session number the recordings $(1-6)$ are derived from and the intended stressfulness of the particular session, coded as an ordinal value $(1-3)$.

Table 15.1 Features extracted from the patient's PTSD profile, reported stress, sessions, and events. The table also indicates their symbols, data types and values

Feature	Symbol	Type and values
Patient profile features *Drawn from patient anamnesis*		
Age	Age	Interval, integer
PCL score	PCL	Interval, integer
Number of deployments experienced	N_{dep}	Interval, integer
Number of days since deployment	N_{day}	Interval, integer
Session features *Derived from game log*		
Session number of the patient (1–6)	S_n	Ordinal, integer
Intended stress of session (1–3)	S_{stress}	Ordinal, integer
Behavioral features *Derived from game log*		
Event type	E_{Type}	Nominal, binary vector
Event number	E_n	Ordinal, integer
Player position X	E_{PosX}	Interval, float
Player position Z	E_{PosZ}	Interval, float
Distance to closest NPC at event time	E_{NPC}	Ratio, float
Time elapsed in session	E_t	Ratio, float
Self-reported stress features *Self-reported by player*		
Patient's rating of session stressfulness	P_{stress}	Interval, integer
Patient's indication of most stressful event	P_{event}	Nominal, binary
SC stress-response features *Derived from continuous decomposition analysis* [3]		
Event response activation	E_{phasic}	Ratio, float
Event response latency	E_{lat}	Ratio, float
Event responses in window	E_N	Interval, float
Tonic SC across window	E_{tonic}	Ratio, float
SC mean in event window	E_{SCmean}	Ratio, float
Maximum deflection in event window	$E_{deflect}$	Ratio, float

Table 15.2 PTSD profile feature values of the subjects who participated in the StarteMart experiment

Feature	Average	Standard deviation	Range
Age	26.8	2.5	22–32
PCL	58.0	4.9	50–65
N_{dep}	1.8	0.7	1–3
N_{day}	1001	432	113–1685

Behavioral Features

For each of the predefined stimulus events taking place in the game, the following features are recorded: The *event-type*, E_{Type} and the *order* of the event in the total session, E_n, the time elapsed since the beginning of the session, E_t, the location of the player's avatar in the virtual supermarket (E_{PosX}, E_{PosY}), and the distance to the nearest non-player character (NPC) in the supermarket, E_{NPC}.

Self-Reported Stress Response Features

Before, immediately after, and following a short break after each of the three sessions, the patient is asked to provide a rating of his subjectively experienced stress level, P_{stress}, on the *Subjective Units of Distress Scale* (SUDS) [32] in a range from 0 to 100 with 0 representing complete absence of stress and 100 representing the most stressful experience the patient can recall. Additionally, the patient is asked to indicate which singular event, P_{event}, during the session was considered the most stressful, if any. This self-report is subsequently reduced to one of the four event categories: social, sound, pickup, and flashback or set to null if no event was indicated.

SC Stress Response Features

Typical trough-to-peak analysis of skin conductivity response (SCR) amplitude, area or similar measures, can be subject to super-positioning of phasic and tonic activity. This may necessitate the subtraction of baseline measures or other forms of signal correction [6]. It has been suggested that even with such corrections one may still confound phasic and tonic SC which is undesirable in a study focusing predominantly on event-related activation [3].

   In order to address this potential issue features of the player's skin conductivity at the time of the event are extracted using Continuous Decomposition Analysis (CDA) as described in [3]. The method allows for the deconvolution of phasic and tonic electrodermal activity. It initially separates superpositioned phasic and tonic components of the raw SC signal. Subsequently it adapts a general model of the human skin's impulse response function (IRF) to the phasic activity by sampling the tonic component around the event response to establish a local baseline and fitting the IRF to the shape of the phasic component. The result is expressed in a *phasic driver* measured in μS that approximates the phasic response affecting the signal within the event window. As such, the phasic driver across the event window can be interpreted as a locally baseline-corrected measure of the patient's SC response to the event. As a result of the deconvolution procedure the phasic driver can take on
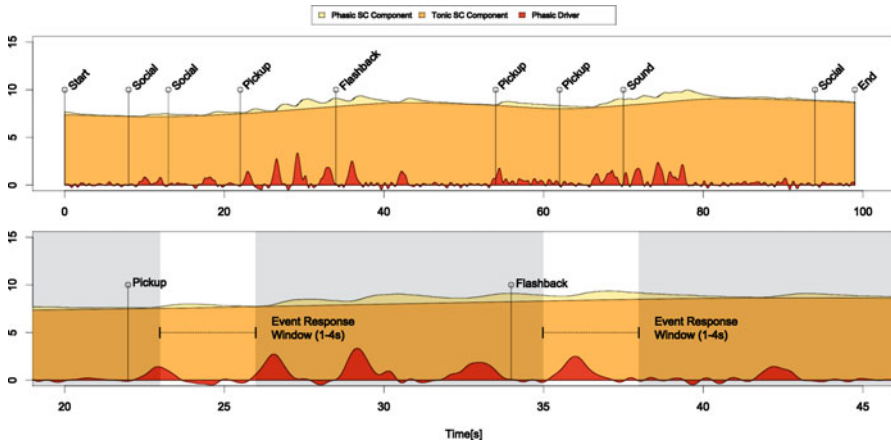
Fig. 15.2 Continuous deconvolution analysis of Player 5, Session 3. The *top figure* shows the full session from the beginning to the end. The *bottom figure* shows a detailed view of an excerpt from the same session. Both figures show three components extracted from the raw SC signal: phasic activity (*yellow*), tonic activity (*orange*), and the phasic driver (*red*) of the SCRs

negative values.[1] A detailed example from the CDA process is provided in Fig. 15.2. More details about the CDA method can be found in [3].

A $1 - 4\,s$ after-event response window is applied, meaning that only activation occurring with this window is considered relevant to the event (see Fig. 15.2). A minimum phasic driver threshold value of $0.05\,\mu S$ is used, meaning that only events with a phasic driver value exceeding this threshold are considered significant and counted as SCRs. From this procedure the following features are extracted: The mean event response activation amplitude across the event window, E_{phasic}, the number of significant SCRs within the response window, E_N, the latency of the first significant SCR, E_{lat}, the mean tonic SC within the response window, E_{tonic}, the mean SC within the response window, E_{SCmean}, and the maximum positive SC deflection within the response window, $E_{deflect}$—see Fig. 15.2 for the detailed description of all SCR extracted features. Combined, these features yield information about the particular event response as well as the general arousal state of the player at the time of the event.

---

[1]A related method that avoids negative phasic driver values, Discrete Decomposition Analysis, has been developed [4], but is generally not recommended by the authors as it is less robust toward artifacts than Continuous Deconvolution Analysis.

Experimental Setup and Protocol

In this section we describe the experimental equipment and configuration used to present StarleMart to the participants and the experimental protocol that was followed for each session.

Equipment and Configuration

The StartleMart Game was presented on an external 25″ flat-screen monitor. Users were placed approximately 35 cm from the screen. Audio stimuli were presented through supra-aural headphones, individually configured to a level that users reported as loud, but not unpleasant. The game was controlled using a standard keyboard (W,A,S,D keys) and mouse, a control scheme typical for first-person perspective computer games.

For recording physiological signals, users were fitted with the Wild Divine IOM biofeedback device. The IOM device records skin conductance (SC) through sensors attached dryly to the little and middle fingers of the user's hand. The values were transmitted to a dedicated recording computer at a rate of 30 Hz via USB. The clock of the recording computer was synchronized to the clock of the computer running *StartleMart* using a local network time protocol server to ensure that events in *StartleMart* could be accurately mapped to physiological responses.

Experimental Protocol

All patients who took part in the study were invited to participate in two separate experimental trials which were scheduled to take place approximately 14 days apart. Each trial included three unique sessions with *StartleMart*. Trials were conducted in a dedicated room at the offices of the patients' regular psychiatrist. Two experimenters, trained psychologists, participated in each trial. One experimenter would mainly interact with the patient, while the second experimenter mainly monitored and managed the experimental setup. The experimenters would welcome the patients and inform them of the nature and purpose of the study. Following the introduction, the patient would complete the diagnostic interview. The experimenters collected background data from the patient and ensured that the patient still met the inclusion criteria for the study. Subsequently, the patient was introduced to the experimental setup and the experimenters would ensure that the patient was familiar with controlling first-person perspective computer games. The measurement device would be attached to fingers of the patient's keyboard operating hand, and the patient would be allowed to familiarize himself with the game controls.

From here, three sessions were completed each following the same template: Firstly, baseline physiological data was collected from the patient. Secondly, the

patient was asked to rate his subjective experience of stress. Thirdly, the patient played through the session in *StartleMart*. Fourthly, immediately after completing the session, the patient was again asked to subjectively report his experienced stress level. Fifthly, the patient used a four-alternative-forced-choice interface to rank all in-game events of the session in terms of stressfulness. Sixthly, the patient reported his subjective experience of stress a third time, before moving on to the next session if any remained. After three sessions, the experimenters debriefed the patient, ending the session.

Results

In order to study the complex relationship between the stress perceived by PTSD patients, the severity of their symptoms and their physiological responses to events in the StartleMart game, we use Multi-Layer Perceptrons (MLPs) to draw the mapping from event context, session and physiological recordings to PCL scores, indicating symptom severity. The prediction accuracy of an MLP inform us to which extent a set of input features is related to the target output while an analysis of its structure and parameters (topology and weights) can reveal the form of this relation. MLPs are versatile tools that allow us to use interval, nominal and ordinal target outputs which are represented by the three types of reports in our dataset. In addition, previous work has demonstrated the usefulness of MLPs of varying complexity in constructing models of affective states [18, 29, 34, 35] motivating their use in this study.

The training of the multi-layer perceptrons is preceded by a model-specific combination of automatic and manual feature selection (FS) in order to eliminate features that are not relevant for the prediction of the target output. We use *backpropagation* (BP) [26] to train the weights of the MLPs. Backpropagation may suffer when the distribution of target outputs is unbalanced. For this reason we resolve class imbalances in the prediction of the most stressful event by randomly under-sampling the dominant class while including all instances of the minority class to achieve an equal number of instances. These experiments were performed using the machine learning toolkit WEKA.[2]

Automatic feature selection (FS) is a standard procedure in data mining and machine learning studies used to remove those input features that are not relevant for the prediction, and thus facilitating the next model training phase. For all experiments we use greedy hill-climbing algorithms that start with an empty set of features and sequentially add the best performing feature to the set until no improvement is gained. For the experiments with ordinal data, we applied the algorithm on its basic form (sequential forward FS) while for the rest we added the possibility of backtracking up to five levels (sequential forward floating FS).

---

[2]http://www.cs.waikato.ac.nz/ml/weka/

Table 15.3 Features selected for predicting *PCL*, after (manual and automatic) feature selection as described in section "Results"

Feature	Symbol
Intended stress of session (1–3)	S_{stress}
Patient's rating of session stressfulness	P_{stress}
Patient's indication of most stressful event	P_{event}
Event number	E_n
Tonic SC across window	E_{tonic}

The performance of each feature subset is calculated as the prediction accuracy of an MLP trained on that subset. The feature selections for these experiments were run with 10-fold cross-validation. The included features are indicated in Table 15.3.

Predicting PTSD Severity

As mentioned above, the PCL score of a PTSD patient is the de facto standard tool for assessing syndrome severity and for tracking progress during therapy. Thus, the ability to predict the patient's PCL score from interaction with the StartleMart game defines a useful feature for psychiatrists using the game for treatment and diagnosis. Estimating a PCL score from game sessions would prove useful in tracking syndrome development, providing another point of triangulation for the evaluation of the patient's status. To enable the prediction of PTSD symptom severity from the collected features, an MLP with a single output is trained. All session features, behavioral event features, and SCR event features (see Table 15.1) are subjected to sequential forward feature selection in order to find the most appropriate features for the model. Selection and training are not performed on any patient profile features, since the PCL score of the patients was only sampled once during the course of experimentation and including this information would arguably implicate training the model to recognize patients rather than predict the PCL score from event responses.

The automatic feature selection results in the selection of the features E_n, E_{tonic}, S_{stress}, P_{stress}, P_{event}. Notably, only two event-level features are selected, while all other features pertain to the whole session, and of special interest is the fact that the phasic response to the events of the game, E_{phasic}, is not picked by the automatic feature selection, but the tonic SC E_{tonic} is included in the feature set. This may suggest that the PCL construct captures aspects of the patient's general state rather than the patient's typical response pattern to individual stressor events.

Different topologies of MLPs are systematically attempted through experimentation, the best performing one being a single hidden-layer 5-7-1 network. The network is trained on 428 event instances for 10,000 epochs and tested with 10-fold cross-validation. The final model achieves a correlation of 0.91 between actual and predicted values, with a root mean squared error of approximately 2.1 points on the PCL scale (which ranges from 17 to 85 in general, but only 50–65 in this

sample, see section "Patient Profile Features" and Table 15.2), and a root relative squared error of approximately 42.9 % indicating that the model allows for precise predictions of unseen PCL values from the five aforementioned features.

As noted above, the PCL instrument is comprised of 17 questionnaire answer items, each yielding a score between 1 and 5 points, which are summed to provide the PCL score. In lieu of this, we consider a mean error of approximately 2 points to constitute a high degree of precision in predicting the PCL score from the StartleMart features, since a two point difference will make very little difference to a therapist in practice, unless the score is close to a cut-off point for the PTSD diagnosis. Typically, a cut-off score of 50 is used in diagnosing PTSD [30].

Discussion

In this chapter we presented a holistic approach to stress detection and PTSD profiling via a game. The analysis shows that it is possible to approximate a patient's PTSD symptom severity based on features extracted from responses to in-game events. This indicates that game-based environments implementing affective computing can be useful tools for diagnosing PTSD. The finding that phasic SCR does not contribute significantly to the model predicting PCL may seem surprising, as patients with a more severe PTSD theoretically should exhibit stronger responses to stressful stimuli. We speculate that this may owe to the fact that the PCL value is a broad indication of the patient's syndrome state and as such, event-level responses might be too specific to capture this high-level information about the patient's relation to stressful environments, while measures of change over longer time spans—in this case a full session—are better indicators of the patient's general state.

It may seem that using a simulation and physiological measurement equipment is an unpractical and cumbersome method of arriving at a screening value for PTSD symptom severity that can also be obtained through a semi-structured interview and/or a questionnaire. Still, we find it important to consider that tools like StartleMart may potentially allow for PTSD symptom severity screening in settings where trained mental health care professionals are not available, and that some patients might be more inclined to engage with the simulation than to engage with a therapist. In the specific case of soldiers, some may also have an interest in downplaying the severity of their symptoms, since future career opportunities and deployments may hinge upon scoring below certain threshold values. As such, any tool contributing to the process of diagnostic triangulation could potentially be of value to psychiatric practice. The results obtained in the analysis should, however, be read with the caveat that the obtained sample of PTSD patients exhibits a variation of PCL scores ranging only from 50 to 65.

A game-based experimental design such as StartleMart necessarily places a significant amount of agency in the hands of the player, making it challenging to ensure that patients are presented with particular experiences in a predetermined

order or at predetermined times in a manner that is coherent with the universe of the game. Though challenging, this is an aspect of the current protocol that should be addressed and improved in future work.

An obvious extension of the presented work would be to extend the applied methods and generated models to larger samples of PTSD patients, possibly even moving beyond veteran soldiers to a patient sample with a more varied etiology. This would allow us to validate the generalizability of the results which, as touched upon above, is currently unknown with the available data. A challenge would be the recruitment of patients for further studies, as obtaining PTSD patients for suitable for experimental exposure studies is logistically challenging and requires a significant amount of ethical concern. In lieu of this challenge, we consider this study a successful investigation into game-based diagnosis and treatment of PTSD in spite of its inherent limitations.

At the more general level, the study presented here shows that there are significant opportunities in using rich, dynamic, game-based environments that allow for high degrees of involvement and immersion. PTSD patients exhibit physiological stress responses to the virtual environment and subjectively experience challenge and stress. This immersion comes at the price of complicating the paradigm. Where more classical, tightly controlled, stimulus-exposure paradigms allow for a well-known and well-ordered, or randomized etc., presentation of stimuli, a game-based exposure environment makes it inherently more difficult to make assumptions about what the patient is experiencing at any given time. The richness of the simulated environment adds to this challenge, making it quite possible for multiple events to occur simultaneously or for the patient to attribute significant meaning to content that was not intended as a stimulus. One option for addressing this issue in future studies could be to construct more restricted game-based environments, returning more control over the exposure situation to the hands of the experimenter or therapist. Whether this would reduce the immersion of the patient is an open question and defines an interesting research question in itself. A second option would be to reverse a central aspect of the paradigm and eschewing the idea of conducting analyses based on predefined events. Instead, it might be possible to identify significant events from the physiological signal, self-reports, and the simulation log. With such an approach it would seem likely that one could obtain a higher number of events and a higher variation in the types of events though the approach would entail a significant effort in terms of data mining and treatment.

Conclusion

In this chapter, we have presented a set of key findings on PTSD profiling and stress detection via games from a sample of 13 clinical PTSD patients interacting with the StartleMart game. Previous work has demonstrated that this game has

an ability to instigate and detect stress in patients to a degree that scales with PTSD symptom severity [12]. The work presented here develops methods for predicting patient symptom severity and subjective experience from behavioral data, physiological responses and game logging data through feature selection and multilayer perceptrons. Using these methods we demonstrate that it is possible to predict patient symptom severity on validated scales with an average deviation of 2 % on the range of the PCL scale.

These findings demonstrate that the approach of using the *StartleMart* game, and by extension games simulating everyday situations, holds potential for psychiatric practice for diagnosing and possible treating PTSD. The game, combined with physiological measurements that can manifest stress responses, provides a tool for obtaining accurate indications of symptom severity from a novel method. The study also demonstrates the challenges of fusing interactive environments with a stimulus exposure approach, since the agency afforded to the player significantly impacts the experimenter's or therapist's control over the flow of events. However, the performance of the computational models of PTSD severity and reported stress derived from the StartleMart dataset indicate that these trade-offs may be useful, since it allows for the construction of an environment that seems to be broadly applicable across patients with variant syndrome manifestations.

References

1. Albright G, Goldman R, Shockley KM, McDevitt F, Akabas S (2012) Using an avatar-based simulation to train families to motivate veterans with post-deployment stress to seek help at the VA. Games Health Res Dev Clin Appl 1(1):21–28
2. American Psychiatric Association (2000) Diagnostic and statistical manual of mental disorders. American Psychiatric Association, Washington, DC
3. Benedek M, Kaernbach C (2010) A continuous measure of phasic electrodermal activity. J Neurosci Methods 190(1):80–91
4. Benedek M, Kaernbach C (2010) Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology 47(4):647–658
5. Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA (1996) Psychometric properties of the PTSD checklist (PCL). Behav Res Ther 34(8):669–673
6. Boucsein W (2011) Electrodermal Activity. Springer, New York
7. Cukor J, Spitalnick J, Difede J, Rizzo A, Rothbaum BO (2009) Emerging treatments for PTSD. Clin Psychol Rev 29(8):715–726
8. First MB, Spitzer RL, Gibbon M, Williams JBW (2002) Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition. (SCID-I/P). Biometrics Research, New York State Psychiatric Institute, New York, Nov 2002
9. Foa EB, Keane TM, Friedman MJ, Cohen JA (2009) Effective treatments for PTSD, Practice guidelines from the international society for traumatic stress studies, 2nd edn. Guilford Press, New York

10. Gates MA, Holowka DW, Vasterling JJ, Keane TM, Marx BP, Rosen RC (2012) Posttraumatic stress disorder in veterans and military personnel: epidemiology, screening, and case recognition. Psycholog Serv 9(4):361
11. Holmes EA, James EL, Coode-Bate T, Deeprose C (2009) Can playing the computer game "tetris" reduce the build-up of flashbacks for trauma? A proposal from cognitive science. PLoS One 4(1):e4153
12. Holmgard C, Yannakakis GN, Karstoft K-I, Andersen HS (2013) Stress detection for PTSD via the startlemart game. In: 2013 Humaine association conference on affective computing and intelligent interaction (ACII). IEEE, pp 523–528
13. Hoque ME, Lane JK, El Kaliouby R, Goodwin M, Picard RW (2009) Exploring speech therapy games with children on the autism spectrum. In: 10th annual conference of the international speech communication association, INTERSPEECH 2009
14. Kashdan TB, Breen WE, Julian T (2010) Everyday strivings in war veterans with posttraumatic stress disorder: suffering from a hyper-focus on avoidance and emotion regulation. Behav Ther 41(3):350–363
15. Kato PM (2010) Video games in health care: closing the gap. Rev Gen Psychol 14(2):113
16. Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB (1995) Posttraumatic stress disorder in the national comorbidity survey. Arch Gen Psychiatry 52(12):1048–1060
17. Mahan AL, Ressler KJ (2012) Fear conditioning, synaptic plasticity and the amygdala: implications for posttraumatic stress disorder. Trends Neurosci 35(1):24–35
18. Martínez HP, Yannakakis GN (2011) Mining multimodal sequential patterns: a case study on affect detection. In: Proceedings of the 13th international conference on multimodal interfaces. ACM, pp 3–10
19. Parsons TD, Rizzo AA (2008) Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. J Behav Ther Exp Psychiatry 39(3):250–261
20. Perala CH (2007) Galvanic skin response as a measure of soldier stress. Technical report, DTIC Document
21. Pole N (2007) The psychophysiology of posttraumatic stress disorder: a meta-analysis. Psychol Bull 133(5):725
22. Reger GM, Holloway KM, Candy C, Rothbaum BO, Difede JA, Rizzo AA, Gahm GA (2011) Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. J Trauma Stress 24(1):93–96
23. Rizzo A, Reger G, Gahm G, Difede JA, Rothbaum BO (2009) Virtual reality exposure therapy for combat-related PTSD. In: Post-traumatic stress disorder, pp 375–399
24. Rizzo A, Buckwalter JG, John B, Newman B, Parsons T, Kenny P, Williams J (2011) Strive: stress resilience in virtual environments: a pre-deployment VR system for training emotional coping skills and assessing chronic and acute stress responses. Stud Health Technol Inform 173:379–385
25. Rothbaum BO, Rizzo A, Difede J et al (2010) Virtual reality exposure therapy for combat-related posttraumatic stress disorder. Ann N Y Acad Sci 1208(1):126–132
26. Rumelhart DE (1995) Backpropagation: theory, architectures, and applications. Lawrence Erlbaum, Hillsdale
27. Schnurr PP, Lunney CA, Sengupta A, Waelde LC (2003) A descriptive analysis of PTSD chronicity in Vietnam veterans. J Trauma Stress 16(6):545–553
28. Siepmann M, Grossmann J, Mück-Weymann M, Kirch W (2003) Effects of sertraline on autonomic and cognitive functions in healthy volunteers. Psychopharmacology 168(3):293–298
29. Wagner J, Kim J, André E (2005) From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification. In: IEEE international conference on multimedia and expo. IEEE, pp 940–943
30. Weathers FW, Litz BT, Herman DS, Huska JA, Keane TM et al (1993) The PTSD checklist (PCL): reliability, validity, and diagnostic utility. In: Annual meeting of the international society for traumatic stress studies, San Antonio

31. Wiederhold BK, Wiederhold MD (2008) Virtual reality for posttraumatic stress disorder and stress inoculation training. J Cyberther Rehabil 1(1):23–35
32. Wolpe J (1973) The practice of behavior therapy. Pergamon Press, New York
33. Wood DP, Webb-Murphy J, McLay RN, Wiederhold BK, Spira JL, Johnston S, Koffman RL, Wiederhold MD, Pyne J et al (2011) Reality graded exposure therapy with physiological monitoring for the treatment of combat related post traumatic stress disorder: a pilot study. Stud Health Technol Inform 163:696
34. Yannakakis GN, Hallam J (2008) Entertainment modeling through physiology in physical play. Int J Hum-Comput Stud 66(10):741–755
35. Yannakakis GN, Martínez HP, Jhala A (2010) Towards affective camera control in games. User Model User-Adapt Interact 20(4):313–340

Chapter 16
Understanding and Designing for Conflict Learning Through Games

Rilla Khaled, Asimina Vasalou, and Richard Joiner

Abstract Conflict resolution skills are fundamental to navigating daily social life, yet means to learn constructive conflict resolution skills are limited. In this chapter, we describe *Village Voices*, a multiplayer serious game we designed that supports children in learning and experimenting with conflict resolution approaches. Drawing on experiential learning as an underlying learning philosophy, and based on Bodine and Crawford's six-phase model of resolving conflict, *Village Voices* puts players in the role of interdependent villagers who need to work their way through conflicts and quests that arise in the game world. In this chapter, we first present *Village Voices* through the design qualities of *competitive collaboration*, *local familiar multiplayer*, *playing around the game*, *reimagining the real*, and *persistence*. We then present a case study that examines the learning experiences of players over four weeks, focusing on the role of time, emotion, the relationship between in-game conflict and learning, and requirements for learning moments.

Introduction

Conflict resolution skills are fundamental to navigating daily social life, but many of us acquire them only piecemeal and indirectly, over a lifetime of social interactions with others. In this chapter, we describe *Village Voices* [19], a multiplayer serious game designed to support children in learning the social skills necessary to constructively engage in conflict resolution. Drawing on experiential learning as an underlying learning philosophy, and based on Bodine and Crawford's six-phase model of resolving conflict [3], *Village Voices* puts players in the role

---

R. Khaled (✉)
Concordia University, Montreal, QC, Canada
e-mail: rilla.khaled@concordia.ca

A. Vasalou
Institute of Education, London, UK
e-mail: a.vasalou@ioe.ac.uk

R. Joiner
University of Bath, Bath, UK
e-mail: r.joiner@bath.ac.uk

of interdependent villagers who need to work their way through conflicts and quests that arise in the game world. In this chapter, we first present *Village Voices* through the design qualities of *competitive collaboration*, *local familiar multiplayer*, *playing around the game*, *reimagining the real*, and *persistence*. We then discuss a case study drawing on examples of game play to show how the game developed children's social emotional skills.

Conflict Education and Games

Our work concerns expressing and learning about conflicts through game experiences. We understand conflict as a process that is initiated when two or more parties involved in an interaction perceive that one member shows or feels strong opposition to the interaction [27]. Such an opposition can arise in relation to parties having different goals. For example, Deutsch states that "a conflict of interests occurs when the actions of one person attempting to reach his or her goals prevent, block, or interfere with the actions of another person attempting to reach his or her goals" [5].

Bodine and Crawford, amongst the most influential voices in conflict resolution education for children, propose the following six-phase process to facilitate young people in dealing with their conflicts proactively and effectively: *Setting the stage:* Assuring students that they will be listened to and not judged, and that all the parties are equally valued; *Gathering perspectives:* Collecting as many points of view as possible; *Identifying interests:* Using communication abilities to determine the underlying sources of conflict, and to focus on people's interests rather than positions; *Creating options:* Using creative-thinking abilities to come up with imaginative, mutual gain solutions to conflict-related problems; *Evaluating options:* Using critical-thinking abilities to apply objective criteria for determining the suitability of a conflict resolution option; and, finally, *Generating agreement:* Coordinating integrated deployment between the two opposing parties, across all of the foundation abilities [3]. Crucially, Bodine and Crawford's process puts young people in the position of resolving their own conflicts.

In the early phases of our user research, we observed that little of Bodine and Crawford's advice was being put into practice in schools. Disputes tended to be settled by third party mediators, responses only came into effect once conflicts had reached critical points, and the breaking of school rules served as a key focal point for determining when a conflict was taking place. Additionally, we learned that many schools have no specific conflict education, and teachers often have no guidance on how to approach conflict resolution in the classroom [28]. This highlighted the potential contribution digital games could make in this space.

Games for conflict resolution have been the topic of previous research. In *FearNOT!*, which focuses on bullying, the player is an invisible friend of a non-player character (NPC) who is a victim of bullying. The player's role concerns advising the NPC on how to cope with bullying-related problems [1]. *Choices*

and Voices is a role-playing game in which players experiment with peer pressure management and resistance strategies, decision making in moral dilemmas, and critical assessment of advice [20]. The interactive scenarios are integrated into a narrative, where players must make a range of decisions and consider different points of view. *Quandary* is a digital card game that presents ethical issues and conflicts involving NPCs for the player to reason through from a mediator perspective, requiring critical thinking, perspective taking, and decision making [7].

In evaluating these existing serious games and how they approached teaching conflict resolution, we identified a number of shortcomings. First, while they included other characters, they were all single-player games, thus not requiring players to deal with other players in exploring and resolving emergent, shared conflicts. Conversely, best practice conflict resolution training frequently puts participants in activities involving other participants because it requires them to practice how to resolve conflicts with other (real) individuals. Second, the games placed players in an advisory or mediator role. On the one hand, this relieved players of encountering the effects of conflict directly, and invited them to approach decision-making in a more objective manner. On the other hand, it did not place players in situations in which they experienced conflict. Formulating and enacting conflict resolution behaviours when personal stakes are involved is considerably more complex than theoretically knowing possible desirable response behaviours. Finally, these games presented conflicts that had been pre-established and set during the game design phase. There was no adaptation to a particular player's sense of conflict: whether a player was experiencing low or high conflict, the game progressed the same way for all. In the next section, we examine the design of our game in light of the aforementioned design opportunities and the design qualities they give rise to.

Village Voices

Village Voices is a four-player open world game that takes place in a fictional village set in a pre-industrial society. It is designed to be played in a classroom setting by players who know one another, under teacher supervision. On the surface, the game is about survival and prosperity in the village. On closer inspection, however, the game is about friendship and reputation management in the village, and mastery of conflict resolution. The game makes use of player models that drive the adjustment and selection of quests for each player. As such, it provides a personalized learning experience for its players, providing them with game quests appropriate to their conflict resolution abilities. Here we examine how *competitive collaboration*, *local familiar multiplayer*, *learning around the game*, *reimagining the real*, and *persistence* inform the design of our game.

Competitive Collaboration

By *competitive collaboration*, we mean game dynamics that invite both competitive and collaborative strategies amongst players. While competition is synonymous with how many people think of games, mixes of competitive and collaborative game mechanics are not unusual in mainstream multiplayer games [30]. Notably, Garzotto points out that in the context of children's learning games, such mixes increase player motivation [9]. With regards to collaborative game mechanics specifically, El-Nasr et al. observe that common player behaviours include helping each other, working out joint strategies, and waiting for each other [23].

The competitive-collaborative combination resonates both with how conflict is understood and how conflict resolution skills are taught. Definitions of conflict, such as the ones we provided earlier, typically foreground the notion that involved parties pursue goals in ways that interfere with those of others [5, 27]. As such, competition is intrinsically connected to conflict, as it too concerns preventing others from obtaining goals. At the same time, best practice conflict resolution approaches advocate solution seeking that satisfies all parties, i.e. collaborative action [3].

We leveraged this mix in designing *Village Voices*, in order to highlight tensions between independent and interdependent goals. As part of daily life in the village, players are required to undertake various actions related to maintenance of their characters' livelihoods (see Fig. 16.1). For example, the alchemist character must tend to his crop of magic mushrooms, and collect and eat fruits to stay healthy.

Fig. 16.1 The Blacksmith in the process of mining for metal (Image used with permission from the FP7 Siren consortium [29])

Players must also complete quests related to their responsibilities within the village. Continuing the example, the alchemist may be in the process of collecting and processing items to build a wall to keep wolves out of the village, which involves trading with other characters. All the while, situations inevitably arise that trigger conflicts or exacerbate existing ones. For example, in order to complete the barrier wall, the alchemist may need to obtain an item from the innkeeper, who he is not on good terms with due to a previous theft incident involving the innkeeper helping herself to the alchemist's mushrooms. While players may initially be faced with simple quests involving no trades or only one trade with other characters, more difficult quests involve trades with all three of the other characters.

Importantly, the game itself does not preach to players about particular ways to behave. Indeed, the mix of competitive and collaborative goals problematises the notion that there is always a correct way to resolve a conflict.

Local Familiar Multiplayer

By *local familiar multiplayer*, we mean multiplayer games that are played by co-located players who have existing histories of social dynamics with one another. As various games scholars have noted, in learning contexts multiplayer games can lead to rich, meaningful, emotionally-charged, and memorable game experiences [6, 16, 24]. Also within the education literature, voices have argued for the merits of co-located peer learning [18], even if it involves arguments [22].

For reasons described above, we had decided to pursue a multiplayer game design. A concern raised by one of the authors, however, was that a multi-player design might lead to situations in which previous or ongoing conflicts between individuals could be exacerbated. Realising that there was no way to prevent this from happening, instead we embraced it as a design dynamic, and considered how we could safely expose existing uncomfortable social dynamics and transform them into learning opportunities.

Village Voices allows players to exhibit destructive as well as constructive behaviours. For example, it is possible to steal from other players and spread rumours about them, as well as give gifts. After each significant interaction with another player, players are asked to update their current feelings towards that player, as well as to gauge the current level of conflict they are feeling (see Fig. 16.2). This acclimatises them towards introspection and self-awareness of their own emotional states but also informs inform how quests are chosen. As players demonstrate progressive competence with regards to resolving conflicts, they are presented with increasingly complex quests relying on negotiation with other players with whom they might have traditionally had problematic relationships offline. We hypothesized that resolving conflicts in-game with known players would be dramatically more memorable and meaningful as learning experiences than those with unknown players, due to their heightened emotional importance and relevance.

Fig. 16.2 The Blacksmith player being asked to gauge the current level of conflict (Image used with permission from the FP7 Siren consortium [29])

Learning Around the Game

By *learning around the game*, we suggest that learning does not only take place while players are engrossed in a game world, or even while game software is running. Learning can take place around play, for example, as a result of reflection or conversation after play. In fact, in the context of simulation gaming, Crookall and Hofstede et al. point out that post-game debriefs, discussion sessions in which the learning implications of games are explicitly addressed, are an essential (and often overlooked) component of unpacking, contextualising, and making sense of simulation gaming experiences [4, 15]. Games with debrief stages are generally more effective in terms of learning transfer than games that lack such a stage.

In their six phase conflict resolution process, Bodine and Crawford place much emphasis on critical thinking and communication skills with either one or both required during the phases of *gathering perspectives, identifying interests, creating options, evaluating options*, and *generating agreement* [3]. But it is currently hard to intuit when critical thinking is taking place via game actions, and developing in-game communication systems that are as expressive as spoken communication remains an unsolved research problem.

Working within these limitations, as well as our research findings on the typical infrastructure of social and emotional learning lessons at school, we designed *Village Voices* play sessions to last approximately 30 min and to take place once a

week over a series of weeks. In each 30 min session, active "play time" was designed to last around 15 min, coinciding with the approximate length of time required to complete one quest. The remainder of the time was set aside for a debrief session to be led by a learning instructor. During the debrief, events that had taken place during play would be reviewed and discussed. Players would be invited to address issues raised, relate game experiences back to life experiences, reason through and advise one another on alternative resolution strategies, and collectively negotiate rules to guide future play sessions.

Needless to say, such a design places an onus on learning instructors to be actively engaged during play sessions and to be ready to unpack play behaviours in post-play contexts. While this may run counter to the vision of "learning in a box" that many teachers and parents have of how learning games function [6], it is reflective of contemporary theories of learning such as situated cognition [25] and distributed cognition [17], as well as ecological approaches to perception and learning [12]. In these paradigms, learning takes place as a result of rich social interaction, and media such as games are understood as tools for learning that exist within broader ecosystems.

Reimagining the Real

By *reimagining the real*, we mean recontextualising real world events within the game world by way of game mechanics, events, narratives, and language. As the literature on learning transfer shows, it is crucial that contextual similarities exist between learned content and application context for connections to be forged and learning to be transferred [8, 26]. At the same time, the idea of creating a school simulation was not appealing from an engagement perspective. In addition, the psychology literature on *minimally counterintuitive concepts* – concepts that are counterintuitive but not too much so – posits that they are easier to recall and repeat than either completely intuitive or completely counterintuitive concepts [2, 13]. Furthermore, Gee argues that what makes games a potentially powerful medium for learning is that they situate meaning in worlds of experience, associating meaning with actions, experiences, images, and dialogue [10].

We dealt with this in *Village Voices* by establishing a game world that was, in some ways, though not all, isomorphic to the school experience. In the *Village Voices* world, characters must act both independently and interdependently in order to survive and not be ostracized from the rest of the player community. Beyond this, however, we modelled many of the game's mechanics on events that we observed taking place at school during the user research phase. The most common types of conflicts we observed concerned accidental harm and jokes gone wrong, deception, friendship, and property disputes [28]. Accordingly, in *Village Voices* it is possible to destroy other characters' dwellings, to deceive others about trades or rumours, to "gang up" on others and also register dissatisfaction with them, and to steal from

others or eat food from their land. Likewise, we made it possible for players to respond in ways that we observed at school. As such, it is possible to be physically aggressive in-game by way of property damage as well as verbally aggressive in-game by spreading rumours and demanding particular friendship allegiances of other players. It is also possible to avoid other players by refusing to trade or collaborate with them.

In recontextualising familiar events within a fantasy medieval setting in which survival is the key objective, we created a minimally counterintuitive but situated learning experience that retained a shared context with daily school life.

Persistence

By *persistence*, we mean games with persistent state that continuously track and respond to player action, and are intended to be played over long periods of time. Serious games are frequently described as "safe environments" for risk-free exploration of behaviours [11, 21]. But an unintentional corollary can be that in-game behaviours have no consequences [14]. While in the context of conflict resolution learning, it is important that players are afforded some degree of safety [3], at the same time, we were wary of creating an environment that felt safe at the expense of feeling important.

We struck a balance by making *Village Voices* a persistent game world in which the same group of four players co-exist together in the village for weeks, and are required to continuously provide updates about their relationships with other players. In this way, both constructive and destructive in-game behaviours have lasting consequences, and players are reminded of their impact on the social climate.

Another motivation underpinning our use of persistence was to invite players to reflect on, and revisit their behaviours. Given that *Village Voices* is designed to be played over a number of weeks, we hoped that the lapses in time between sessions coupled with debriefs would give players a chance to observe their own behaviour from afar. Furthermore, we envisioned that players might plan strategies with other players, and even establish codes of conduct for everyone to follow.

Case Study

The *Village Voices* game was played by five groups in the UK, each with four children (a total of 20 children). In this chapter we present findings from one of those groups, composed of four late primary school students (three boys and one girl) between the ages of 9–10. Students played the game once a week for a period of 4 weeks in a private room at their school. Each session lasted approximately

60 min, of which 10 min were dedicated towards setup, 40 min of which involved gameplay, and the remaining 10 min of which were used for a debrief conversation. Students were seated across from each other such that they could engage in active dialogue during gameplay. Each play session was video recorded. At the end of each session, a research assistant who had been present as an observer conducted a post-game reflection. Our analysis is inductive, and draws on the dialogue during and around gameplay to understand the impact of the game on children's conflict resolution skills.

Conflict Experiences and Skills Become More Nuanced Over Time

Over the 4 week period, players' experience of conflict in the game evolved in pace with their involvement in it. During the first week, whenever players engaged in a conflict, they expressed detached amusement about their actions. By the second week, however, and until the end of the study, players were demonstrating emotional investment in their characters and had clearly defined relationships with other player characters in the game. As a consequence, they often became emotional about the conflicts they experienced with others, arguing during game play when they perceived themselves to be in the role of the victim. Initially, players defaulted to employing competitive strategies with one another. As the weeks passed, collaboration became more frequent as a strategy. Notably, it never fully replaced competition, but rather became progressively intertwined with competitive strategies during and between play sessions.

In-Game Conflicts Do Not Always Engender Learning Moments

During the first session, players began to play the game by applying competitive and collaborative strategies without actively reflecting on their choices. On application of the strategies, however, players did not necessarily pause to reflect on the consequences they engendered. For example, the innkeeper wanted wood from other players. The carpenter replied to say that he had wood, but before he had a chance to offer a trade, the innkeeper destroyed his house. The carpenter in turn responded, "Why did you destroy my house? That's it, I'm going to get the innkeeper, kill the innkeeper!" The carpenter mirrored the innkeeper's competitive strategy without considering the option of negotiation. Thus we suggest that encountering a conflict during play is a necessary but not sufficient condition in prompting players' understanding of the consequences involved, or the benefit of using more collaborative strategies.

Learning Moments Are Shared, Communicated, and Emotionally Challenging

Collective consequences: As players became more fluent with the game's competitive strategies (e.g. stealing), as a group they were often unable to progress in the game as each player employed a 'tit-for-tat' approach that led to intense episodes of conflict. But on observing the detrimental consequences of competition on the game as a whole, children started to use more collaborative strategies.

Communication skills: By offering a strategy for stealing, *Village Voices* also offered a zero-sum approach to conflict. Players often broadcasted their intention to steal with one another. For example, in one session the carpenter threatened, "Whose house is the innkeeper's house? If you don't tell me I will be stealing from a random house". After repeated exchanges between the carpenter and the innkeeper, the innkeeper communicated a negotiation strategy using game language that would fulfill the carpenter's goals: "Don't steal, don't steal from my house... Don't steal. Trade." He thereby convinced the carpenter to trade so that both parties could gain from the interaction.

Emotion as a learning trigger: During the first 2 weeks of play, the alchemist instigated the majority of conflicts. In the second session, after the alchemist had stolen grain from the innkeeper, the carpenter in turn stole it from the alchemist and offered it back to the innkeeper. This forged an alliance between these two players against the alchemist, which continued for the remaining sessions. Eventually this alliance escalated into a targeted theft from the alchemist's house. Imploring other players stop stealing his resources, the alchemist said, "Can everyone please stop taking stuff from me? Now the innkeeper is taking all of my stuff. This is getting ridiculous. Everybody is taking my stuff. If everybody keeps taking my stuff I'm going to quit the game." Despite the plea, by the end of the session the other players had stolen everything from the alchemist's house and he broke down in tears. The following week, when students returned for another session, they demonstrated that they had reflected on their own house rules, as the game system itself leaves many rules open to debate. Negotiating his boundaries, the alchemist argued that he was willing to accept some competition: "Okay guys, you have 2 min to steal whatever you want from me". In response to this, however, the innkeeper declined on the grounds that such an action might escalate into collective conflict: "No, don't steal or we're all going to end up stealing".

Discussion and Conclusion

Over the course of the 4 week evaluation period, our observations of player behaviour revealed that *Village Voices* invited behaviours from several of the phases advocated for in Bodine and Crawford's approach to conflict education. These included *communication, gathering perspectives, identifying interests, creating options*, and *generating agreement*.

In line with our design quality of *competitive collaboration*, we observed players using both competitive and collaborative actions in responding to conflict, with collaboration being the preferred strategy as the weeks progressed. For players to experience conflict, it was necessary for one or more players to make use of competitive actions. But competitive actions alone were not enough to create a perception of conflict, as players approached their conflicts with a sense of detachment during early play sessions. Instead, perception of conflict was tied to how emotionally committed players were to their characters and their relationships with other player characters. We suggest this arose as a result of the *persistence* inherent in the design of the game, which afforded enough time for players to become acclimated to the game and for character attachments to form, as well as the *local familiar multiplayer* design, which leveraged players' existing social relationships with one another.

Reflection on conflict, however, seemed to be largely prompted when players were able to draw connections between their actions in-game and their consequences, in keeping with *reimagining the real*. We observed this happening both in terms of game progress (e.g. when 'tit-for-tat' approaches stalled progress) as well as player responses to game events (e.g. when players ganging up against another player resulted in the target breaking down in tears). The latter highlighted the importance of *learning around the game* and *persistence*. Between the second and the third sessions, players who had previously been favouring competitive approaches clearly experienced a change of heart and began session three by collectively negotiating acceptable, constructive play strategies, drawing on their previous play experiences to justify their approach. Crucially, the decision to play more collaboratively and less competitively came from the players themselves, which stems from the *local familiar multiplayer* nature of the game.

In closing, in this chapter we have examined the design of *Village Voices*, our multiplayer serious game that supports children in learning and experimenting with conflict resolution approaches. Specifically, we showed how the qualities of *competitive collaboration*, *local familiar multiplayer*, *playing around the game*, *reimagining the real*, and *persistence* informed the game's design. We then presented a case study of the experiences of four student players between the ages of 9–10 over a four week period. As well as showing how the design qualities came to life, we observed that conflict experiences and skills became more nuanced over time, in-game conflicts did not always engender learning moments, and that the most powerful learning moments were shared, communicated, and emotionally challenging. We believe that the design qualities profiled here alongside our qualitative findings can help in the design of serious games beyond the domain of conflict education strengthening the growing and crucial design direction of learning games focused on soft skills, collaborative dialogue, and emotional intelligence.

References

1. Aylett R, Vala M, Sequeira P, Paiva A (2007) Fearnot! – an emergent narrative approach to virtual dramas for anti-bullying education. In: International conference on virtual storytelling, Saint Malo
2. Banerjee K, Haque OS, Spelke ES (2013) Melting lizards and crying mailboxes: children's preferential recall of minimally counterintuitive concepts. Cogn Sci 37(7):1251–1289
3. Bodine R, Crawford D (1998) The handbook of conflict resolution education: a guide to building quality programs in schools. Jossey-Bass Publishers, San Francisco
4. Crookall D (2010) Serious games, debriefing, and simulation/gaming as a discipline. Simul Gaming 41(6):898–920
5. Deutsch M (2006) Introduction. In: Deutsch M, Coleman PT, Marcus EC (eds) The handbook of conflict resolution: theory and practice, 2nd edn. Jossey-Bass Publishers, San Francisco, pp 1–22
6. Egenfeldt-Nielsen S (2007) Third generation educational use of computer games. J Educ Multimed Hypermed 16(3):263–281
7. FableVision (2012) Quandary. http://www.quandarygame.org/
8. Fisch S, Kirkorian H, Anderson D (2005) Transfer of learning in informal education: the case of television. In: Mestre JP (ed) Transfer of learning from a modern multidisciplinary perspective. Information Age Publishing, Greenwich, pp 371–393
9. Garzotto F (2007) Investigating the educational effectiveness of multiplayer online games for children. In: Proceedings of the 6th international conference on interaction design and children, IDC '07. ACM, New York, pp 29–36
10. Gee JP (2009) A situated sociocultural approach to literacy and technology. In: The new literacies: multiple perspectives on research and practice. The Guilford Press, New York, Chap 8
11. Geurts J, de Caluwé L, Stoppelenburg A (2000) Changing organisations with gaming/simulations. Elsevier Bedrijfsinformatie, 's-Gravenhage
12. Gibson E, Pick A (2000) An ecological approach to perceptual learning and development. Oxford University Press, Oxford/New York
13. Gonce, Lauren O, Upal, Afzal M, Slone, Jason D, Tweney, Ryan D (2006) Role of context in the recall of counterintuitive concepts. J Cogn Cult 6(3–4):521–547
14. Hijmans E, Peters V, van de Westelaken M, Heldens J, van Gils A (2009) Encounters of a safe environment in simulation games. In: Bagdonas E, Patasiene I (eds) Games: virtual worlds and reality. Selected papers of ISAGA 2009
15. Hofstede GJ, De Caluwé L, Peters V (2010) Why simulation games work-in search of the active substance: a synthesis. Simul Gaming 41(6):824–843
16. Hromek R, Roffey S (2009) Promoting social and emotional learning with games: "it's fun and we learn things". Simul Gaming 40(5):626–644
17. Hutchins E (1995) Cognition in the wild. A Bradford book. MIT Press, Cambridge
18. Johnson DW, Johnson RT (1989) Cooperation and competition: theory and research. Interaction Book Company, Edina
19. Khaled R, Yannakakis GN (2013) Village voices: an adaptive game for conflict resolution. In: Proceedings of the 8th international conference on foundations of digital games, pp 425–426
20. PlayGen (2010) Choices and voices. http://playgen.com/choices-and-voices/
21. Raybourn EM (2000) Designing an emergent culture of negotiation in collaborative virtual communities: the case of the domecitymoo. SIGGROUP Bull 21(1):28–29
22. Schwarz BB, Neuman Y, Biezuner S (2000) Two wrongs may make a right . . . if they argue together! Cogn Instr 18(4):461–494
23. Seif El-Nasr M, Aghabeigi B, Milam D, Erfani M, Lameman B, Maygoli H, Mah S (2010) Understanding and evaluating cooperative games. In: Proceedings of the SIGCHI conference on human factors in computing systems, CHI '10. ACM, New York, pp 253–262

24. Squire K (2006) From content to context: videogames as designed experience. Educ Res 35(8):19–29
25. Suchman LA (1987) Plans and situated actions: the problem of human-machine communication. Cambridge University Press, New York
26. Súilleabháin GÓ, Sime JA (2010) Games for learning and learning transfer. In: Donnelly R, O'Rourke KC, Harvey J (ed) Critical design and effective tools for e-learning in higher education: theory into practice. IGI Global Publication, Hershey, pp 113–126
27. Thomas KW (1992) Conflict and conflict management: reflections and update. J Organ Behav 13(3):265–274
28. Vasalou A, Ingram G, Khaled R (2012) User-centered research in the early stages of a learning game. In: Proceedings of the designing interactive systems conference, DIS '12. ACM, New York, pp 116–125
29. Yannakakis GN, Togelius J, Khaled R, Jhala A, Karpouzis K, Paiva A, Vasalou A (2010) Siren: towards adaptive serious games for teaching conflict resolution. In: Proceedings European conference on games-based learning (ECGBL), Copenhagen, pp 412–417
30. Zagal JP, Nussbaum M, Rosas R (2000) A model to support the design of multiplayer games. Presence: Teleoper Virtual Environ 9(5):448–462

Chapter 17
Games Robots Play: Once More, with Feeling

Ruth Aylett

Abstract In this chapter we first examine the requirements for social game-play robots in three game scenario types: robot play companions, robots and digitised games, robots and augmented reality. We consider issues relating to affect recognition, affective modelling in the robot, and robot expressive behaviour. We then discuss work in each of the three scenario types and how it has attempted to meet the requirements advanced. Finally the chapter considers key research issues for the future.

Requirements for Social Game-Playing Robots

Discussion of robots in the context of digital gaming may seem a little paradoxical. After all, the major characteristic of robots is that they are part of the real physical world and not the virtual digital world. So how can they be involved in *digital* games? Indeed, RoboCup football is the most obvious intersection between robots and games. However this is a specialist area and as it involves robot-robot collaboration rather than human-robot interaction, it is not the topic of this chapter.

Three areas of application come to mind. The first and earliest extends the definition of *game* into play, and involves the creation of robot play-companions for children. The second draws on new collaborative display technologies such as multi-touch surfaces. Here, a robot acts in the real world in the roles a human might otherwise occupy, from fellow-player in digitally-supported board games [33], to intelligent tutor in serious games with educational purposes [34]. Finally, if the purely digital is moved into the real world via augmented reality approaches, then one can consider a robot actor that becomes part of an overall game experience containing both real and virtual elements.

All of these roles put the robot firmly into the relatively new research domain of social robotics [13]. A *social robot* can be defined as: "a physical entity embodied in a complex, dynamic, and social environment sufficiently empowered to behave in a manner conducive to its own goals and those of its community" [9]. Clearly

R. Aylett (✉)
MACS, Heriot-Watt University, Edinburgh, Scotland, UK
e-mail: ruth@macs.hw.ac.uk

affect is a fundamental human characteristic in social situations, with a particular impact on non-verbal behaviour. In the human case, the three areas just mentioned could involve for example a human player gloating about the success of their move in a collaborative board game, a child 'telling off' a robot doll for bad behaviour, a human player in an augmented reality scenario expressing encouragement to a robot actor in a shared scenario.

It is considering these concrete scenario types that allows us to derive some generic requirements for game-playing robots. These relate to capturing user affective state, modelling the state of the user on the robot side, and robot expressive behaviour.

All three scenario types suggest that a social game-playing robot requires information about the affective state of the humans with whom they are interacting. Without this information it is very hard to respond appropriately, impacting the success of the robot's action-selection mechanism. Detecting the affective state of a human interaction partner can be viewed as a difficult problem in social signal processing (as discussed in part II of this book), and remains work in progress. The current state-of-the-art [42] is based on multi-modal approaches that try to fuse data from facial expression, voice, and sometimes, physiological data such as skin conductivity. However levels of accuracy on a small set of affective states (four or five) remain well below the 95 % needed and this raises problems for social game-playing robot applications for which specific solutions must then be engineered.

There are theoretical problems as well as signal processing problems in assessing the affective state of a user. The term *state* may itself be one of them, given that human interaction partners move through a dynamic affective process in which affect is typically not very strong and more likely to be irritation than anger, boredom than fear, and amusement rather than outright happiness. A second well-known problem is that mapping between the behaviour captured by the robot's sensors and the affective state of the user is not straightforward. A smile is a relatively easy expression to capture with modern sensors, but smiles are one of the most ambiguous facial expressions [21], able to express embarrassment, sorrow, anger and social welcome or agreement, among other things, as well as happiness.

Fortunately, a game-playing robot has a big advantage over a robot in less structured social situations. This is because the context for interaction is a *game*, with known rules and moves. As we will see later in the chess-playing companion robot, this helps a great deal because it becomes possible to infer affective state with a high likelihood of accuracy. In general, a combination of inference and sensor processing seems a promising approach, especially given the evidence for the human use of expectations in processing interaction data [38].

As well as trying to capture the affect of interaction partners, there are strong arguments for a game-playing robot itself attempting to impact the user's affective state [39]. This may be for dramatic or pedagogical reasons, but it may also be a way of simplifying the sensor-processing problem just discussed. If a robot performs a game-action that is intended to please a human player, and the human then smiles, this makes happiness the most likely interpretation of the smile.

Both inferring user affect from situation and changing user affect via actions require a robot that has a model of affect [1] such that it is able to carry out some aspects of Theory of Mind (ToM) processing [7] about the situation of the human interaction partner. This need not be an explicit model [25] – it could be embedded in a set of behaviour rules or stimulus-action pairs. It could be as specific as a rule that takes the metric of goodness of a chess move and maps that onto the likely happiness of the person that made the move. However there are advantages in an explicit model where the game-playing robot needs to use natural language outputs as part of its action repertoire and where the robot has a more flexible set of roles and functions rather than specialising in one thing, like playing chess. Cognitive appraisal models [31] are useful in this context, especially because running in one direction they can generate robot expressive behaviour, while running in the other, they offer a simulation view of ToM affect issues (e.g. 'What does my model say I would feel if I did that?') [7, 30, 35].

There are a number of reasons for wanting the robot to be able to generate contextually appropriate robot expressive behaviour. One is that communicating the robot's affective state is one way of impacting that of its interaction partner. Thus robot expressive behaviour can be perceived by human players as a commentary on the game-actions of the robot, but also as comment on the game-actions of human players. However there is also a more generic reason for producing expressive behaviour.

Each of the three scenario types above supports a set of game-based social roles both for the robot and its human interaction partners, and each role has an associated set of expectations about behaviour and goals. A robot can signal adherence to its expected role by creating transparency about its current goals and by advertising its coming behaviours before it executes them. One could do this verbally, but in smooth human-human interaction, this is far from the only or even the dominant modality. Taking an *intentional stance* [6] means that in human-human interaction the participants are continuously monitoring each other's goals and updating their expectations about behaviour, largely through the use of non-verbal behaviours such as gaze, facial expression, posture, gesture, non-verbal vocalisations [29].

While the generation of expressive behaviour has been extensively researched for graphical characters [32], and attempts have been made to create standard interfaces supporting expressive mark-up languages [19, 40], social robots raise a set of new issues, primarily due to the very variable physical embodiments they may involve. Indeed, issues relating to embodiment are fundamental throughout HRI research. For a reasonably humanoid robot like the Nao, it may be feasible to apply similar methods as for graphical characters [22]. However robots need not be humanoid – they could be more animal in form, as in the case of the Kismet [3] or Paro [16] robots, or very machine-like in design. There is no absolute requirement that they have facial features (the Nao has almost none) or indeed anything like a face; limbs with which to gesture; or an articulated body supporting posture. If issues of naturalism are moot with graphical humanoid characters, they are much more so in general with robots.

While we have argued that games help with some of the problems of affective processing in social robots, there is at least one area in which they may make these problems more difficult. Games provide a rare example of a social situation in which participants may be entitled to deceive fellow players about their goals and likely behaviour [7]. The card game Poker is an obvious example, but there are many others: board games such as Risk where secret alliances between players may be formed, and Mafia/Werewolf, where the point is to dissimulate if you are a Mafia member/Werewolf [7]. In these cases, generating robot expressive behaviour requires a more indirect mapping between what the affective model says and what the robot expresses. ToM processing is needed for the robot to assess the likely impact of its choice of deceiving behaviour. Assessing the expressive behaviour of human players also becomes a much more difficult problem.

In the rest of this chapter we consider recent work in the three game types just outlined, focusing on the interplay between embodiment and expressive behaviour, and to its relationship to empathic engagement between robot and human interaction partners. We conclude with a look at the most significant continuing research issues.

Robots as Play Companions

The introduction of robots as play companions has come both through the evolution of actual toys in the direction of robot functionality and through the development of research prototypes.

Graphical artefacts such as the Tamagotchi [8] of the later 1990s established that human-like or even sophisticated embodiment is not required to evoke attachment to a play character. The Tamagotchi was a small and very low-resolution graphical artefact on a small plastic key-fob, loosely based on a chicken. Its behaviour appeared autonomous; the human was responsible for feeding it virtual food and giving it caring behaviour through frequent interactions to which it would respond. If neglected or 'old' it would 'die'. Its small behaviour repertoire was wholly expressive of its inner state since there was no functional content, no 'task' that a Tamagotchi had to carry out. These behaviours were enough to create a caring, empathic relationship on the human side and there were anecdotal accounts of children mourning when their Tamagotchi 'died', raising ethical issues that had not been addressed in the design [37].

This was followed in 1999 by the Sony Aibo [14], marketed as a robot pet/adult toy and embodied as a mechanical-looking dog. A design criterion was 'lifelikeness': this could have been trivially addressed by making the appearance more naturalistic, for example by providing a furry exterior. However instead this was interpreted as a behavioural requirement, that of 'maximising the complexity of responses and movements of the robot [14]. This did have embodiment implications. Complex movement requires a higher number of degrees of freedom so that different body parts can be moved in different combinations. Complex responses also require sufficient sensor input. In addition, designers saw the need for multiple motivations

for movement, and mechanisms for producing non-repetitive behaviour so that the Aibo did not become predictable like a normal machine.

Again, almost all Aibo behaviour was expressive, driven by four 'instincts': affection, investigation, exercise, and appetite and expressing the six Ekman 'basic emotions': joy, sadness, anger, fear, surprise and disgust. Emotions allowed different behaviours to be produced in the same external situation because they reflected the impact of earlier interactions and thus altered the context. Thus a happy Aibo would produce a paw extension behaviour when its sensors detected a hand in front of it, while an angry Aibo would not. This underlines the close relationship between an architectural model of emotion in a robot and the generation of a rich behavioural repertoire.

The Aibo was a relatively successful consumer product by the standard of social robots, though never on the scale of Sony's other range of devices, which may explain why it was eventually discontinued. The scale and type of engagement with Aibos was researched by a number of groups, and in [15], analysis of more than 2000 forum postings by Aibo owners suggested a high degree of affective engagement.

Other robot toys have followed, with the Pleo, a small and relatively cheap dinosaur, having some commercial success. However analysis of Pleo in the home brought out quite sharply the current limitations of robots as free-play toys. After the novelty effect had worn off, the Pleo's limitations led to diminishing use over time [11]. These limitations included straightforward issues like short battery life. The Pleo could not act as a toy when it was being recharged because its battery had to be extracted from its belly and put into a charger. The Aibo had a much more sophisticated autonomous charging system in which the robot would move itself into the charger, but this reflected its much higher cost. Even then, it too could not act as a plaything while being charged.

While this is an issue of embodiment rather than robot game-behaviour, it also seems likely that such robot toys still have too little flexibility for a child's open-ended play as well as too little functionality to be able to drive the play interaction themselves for more than a limited period. These limitations motivated the development of the iCat Chess Companion developed in the EU project LIREC [27] between 2008 and 2012.

The iCat [4], designed as a research tool by the Philips Eindhoven laboratory, was what is known as an *interface robot*, designed to rest on a desk and communicate rather than to move around. It took the form of a small yellow plastic cat with programmable head movement, eyes, mouth and eyebrows. The cat had a schematic body but no legs (see Fig. 17.1). It was capable of flexible facial expressions, glance and head movement as well as speech output. Only a limited number were produced and it was never commercially available. The design is not naturalistic and raises no expectations of real cat behaviour.

An iCat was developed into a chess opponent for children in a Portuguese chess club. An advantage of this application is that the role of a chess opponent is well understood and also quite limited. The game is turn-based. Functionally, such a robot needs to be able to play chess, an easy requirement to meet given many

Fig. 17.1 iCat chess
companion

mature chess engines are already available. To avoid the more difficult physical functionality of moving the pieces, the child player was asked to move the piece for the iCat.

However though being able to play chess is the starting point for a chess opponent, as a child's leisure activity it also has a social dimension, with comment about the moves and how each player is doing. Thus it is a tractable environment in which to explore the impact of robot expressive behaviour and the development of an empathic engagement, in which the child player is willing to assign to the robot an ability to understand how they, the child, feel.

The iCat first used the game state reported by the chess engine to adopt a sad or happy expression depending on whether its game position was strong or weak. A study [24] showed that this allowed the child player to more accurately assess the game state than in a version without the happy and sad expressions. Next, an affect detection system was developed so as to assess the engagement of the child player so that the iCat could respond empathically, that is, with actions taking the child's affect into account.

The difficulty of generalised affect detection is not the only reason for developing a scenario-dependant system, as was done in this case. The affective states of interest are themselves scenario dependant [5], concerned both with the content of the interaction (playing chess in this case) and with other interaction parameters such as whether the user is standing or sitting, moving or stationary, one or many, facing or not the social robot. Many affect systems have been developed using actors; in this case the aim was to use in-the-wild data of children actually playing chess.

A person-independent Bayesian learning system was developed using a video corpus of children playing chess, but knowledge from the task and the social context was also included and improved the recognition rates [5]. Smiles, gaze at the robot, game state and game evolution were used to generate probability values for the child's positive, neutral and negative valence at each turn. The iCat would generate an expected next move for the user, and then compare the actual move to this expectation. In the style of cognitive appraisal, this comparison was used to generate an iCat emotion and an accompanying expressive behaviour for it. There were nine of these depending on whether the next move was better or worse than expected [5]: Excited, Confirm, Happy, Arrogant, Think, Shocked, Apologise, Angry, Scared. Some of these – Happy, Angry, Scared – correspond to straightforward emotions, and others – Confirm, Think, Apologise – do not. This underlines the point that affect for a social robot is frequently based around quite complex affective states. Evaluation showed that the child players noticed the empathic elements of the iCat interaction, and had a more positive attitude to it as a result [23]. An interesting aspect was that though the iCat had no speech recognition capabilities, the child players did not remark upon this and it did not seem to impact the interaction.

Robots and Digitised Games

The second area involving robots and digital games identified above generalises the iCat example into a robot player of digital board games. This exploits the growing availability and popularity of large multi-touch surfaces that allow a board game to be ported into a digital form. A characteristic of such games is that they are often multi-player so that social interaction is correspondingly more substantial and more complex than in a two-player game like chess.

Given that full natural language interaction using speech is well beyond the state of the art and one cannot expect a social robot to sustain a conversation, this may seem likely to worsen the problems already discussed. However one should bear in mind that in multiple player games, human-human interaction becomes more important. With two human players and a robot, human-human interaction occupies a third of the possible interactions space. With three human players and a robot it is half of the possible interactions space. Thus the robot is no longer wholly responsible for the social experience.

Moreover multi-person conversation in a game context is often fairly unstructured, with stereotypical speech actions like announcing whose turn it is, and repartee, rather than sustained conversation, around the game state.

Early work on a social robot for a multi-player game took place in the LIREC project [27] in which an EMYS robot head (see Fig. 17.2) was incorporated into a multi-touch surface version of a game called Risk [33]. This is a popular board game for three to six players in which players run armies and try to conquer territories. Successful play involves the creation of informal alliances and joint attacks as well as deception and betrayal.

Fig. 17.2 EMYS robot playing a game of risk

Believable verbal and non-verbal behaviours were seen as key requirements and a robotic embodiment as a way of supporting non-verbal behaviours. As in this article, the ability to model affective state was specified as a requirement, as well as the ability to simulate social roles commonly found in board games. These roles included Helper, a role already discussed for the iCat chess companion, but also Dominator, in which the agent tries to influence a human player to take a specific action. A role not so far discussed was Exhibitionist – behaviour intended to grab attention. Finally, a role specific to a multi-player game was Negotiator – mediating between two other players.

A requirement also present in the iCat chess companion and clearly very significant for any social robot in or out of a game was that of being able to recognise the human players, with an associated 'greet' behaviour, and to remember interactions with them over time [33].

Again, this is in principle a more difficult requirement in a multi-player game than in a single-player one, where the robot has no choice of player with whom to interact. However in a turn-based board game, players normally play in a fixed order and sit in a stable configuration round the board, so sophisticated facial recognition is not needed. The game context comes to the rescue once more. In addition, a microphone array can determine which player is speaking, though in a multi-player game overlapping utterances and interruptions are very common.

This allows a gaze behaviour to be developed so that the game-playing robot can 'look' at the player with whom they are interacting. As the iCat chess companion showed, this is a very powerful social behaviour and is probably one of the factors supporting acceptable social interaction without speech recognition. In a board game, players also often look at active areas of the board, and on a multi-touch surface it is relatively easy for a robot to determine which these are given the use of touch to make a move.

Fig. 17.3 An Aldebaran Nao torso-only robot plays enercities

Fortunately it is a great deal easier to generate speech utterances for a social game-playing robot than it would be to recognise the speech of the other players. The game itself can be used to determine when the robot should speak, and in the work described here, a *relevance* value was used to select significant game events and prevent the robot tediously commenting on everything. Highly relevant events included game attacks on the robot's forces, or conquests by a player the robot *disliked*. Like and dislike were established via alliances and attacks. A further source of utterances was the die roll, where 'good' or 'bad' values often produce human utterances about luck.

This work has been extended into a serious game environment in the EMOTE project [10], in which the sustainable energy game Enercities [17] was converted from a single-user web-based system to a multi-player touch-table based game [34]. The robot is a torso-only version of the Aldebaran Nao (Fig. 17.3).

Where the robot player of Risk was simple a player in a similar role to the human players, in Enercities, the robot player is expected in addition to play a tutorial role. This underlines that a Robot Digital game player may sometimes be more than just a player, and its model of the game may need to be more sophisticated than telling it which is the best move in relation to the game rules. In the Enercities case, there may be cases where the robot should play a non-optimal move from the point of view of game score if this produces good teaching for the human players. The social roles involved are now different from the those of the peer-player robot of Risk: for example an Exhibitionist role becomes wholly inappropriate, while the Negotiator role is folded into a pedagogical approach in which the trade-off between possible player-choices are one of the learning points. Careful analysis of social roles is needed as much as game rule competence for a successful digital game-playing robot.

Robots and Augmented Reality Game Experiences

The vision of Pervasive Gaming is to take digital games out of the computer-based virtual world and into the real physical world using an augmented reality approach [28]. This has so far largely involved the use of handheld devices, but the incorporation of actor-robots seems an interesting and logical step to contemplate. Most initial work in this field has focused on theatre [12, 18, 26], and there are good reasons why this should be so.

A theatrical stage is a much more controlled environment than a pervasive game, which could be room-based but often takes place in the normal physical world. The stage is a privileged environment to which extra resources can be added, for example to give robot actors other information than that directly available to their sensors. In addition, theatrical drama is usually pre-scripted, and while this does not remove the need for real-time responsiveness, interacting with human actors following a script is much less difficult than interacting with improvising pervasive game players. In particular, the problem of speech-based natural language interaction is finessed by knowing in advance what utterances fellow actors will make.

However it is also true that a theatrical performance is required to engage a large audience of spectators, where a pervasive game involves a much smaller number of participants. Participants must themselves take timely actions and are therefore more able to cope with robot actor errors, as long as they are not too frequent and substantial. As we saw in the case of the iCat and digital board game robots, as long as relevant robot utterances are combined with appropriate expressive behavior, especially glance, participants may overlook the absence of natural language interaction. However, spectators are likely to notice robot actor errors very quickly and these are more likely to impact their enjoyment. For this reason, much theatrical robot work involves a greater or lesser degree of tele-operation, especially where a real-world performance is involved [41].

One approach that is argued can help humans to accommodate robot actor errors is clearly signaling the behavior it is about to carry out and then expressing a reaction to its success or failure, much as animators do with graphical characters [18]. Robot actors make a good testbed for exploring this approach since it is common in human drama too, and is an example of how integrating robots into game environments may have a useful spin-off for more general social interaction.

A profound difference between this type of scenario and the ones so far considered is that the robot must navigate. Robot toys may move but they do not have to navigate, and they could, equally, be moved by the child. Robot board-game players should not move away from the board. Robot actors must move according to the logic of the dramatic situation, whether to specific locations, or in order to interact with specific human players. It is possible to design a drama or indeed a game around the known limitations of the robot, but the experience a 2014 performance in Pittsburgh involving the HERB robot revealed that precisely because the robot had wheels and could move, the audience expected it to do so in performance [41].
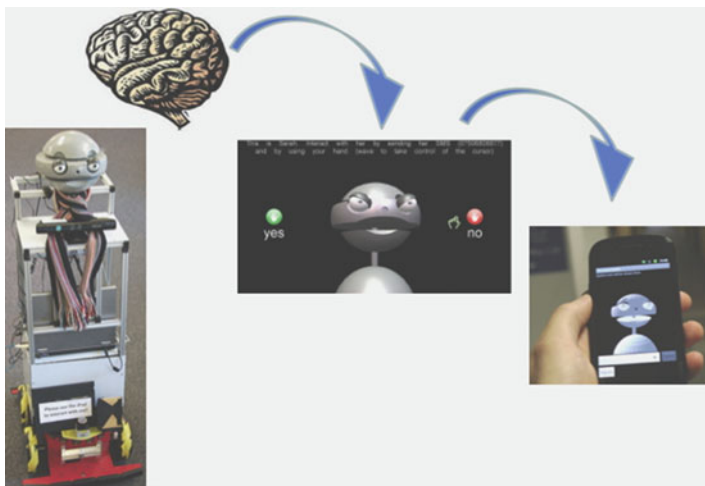
Fig. 17.4 Migration from robot to graphical character

A novel approach top dealing with robot movement constraints explored *migration* – a transfer of the robot 'character' into a graphical embodiment and back [20]. This was applied to a treasure hunt game in which an Emys robot (as in Fig. 17.4) was transferred in the form of small animated version to a handheld device the accompanied the player as they collected clues [2]. This has not been tried in theatre applications, where it seems inappropriate for a static audience, but could be a promising approach for pervasive games.

This work demonstrated that although a robot has greater physical presence than a graphical character [36], a common appearance and voice are strong-enough cues for users to assume a continuity of personality between very different embodiments. A study in which one version of the migration allowed the character to refer to memories across embodiments while another version had the character forget interactions from earlier embodiments, showed that this had no effect at all on the user perception of it being 'the same' character [2].

Key Research Issues

In this article we have looked at three possible ways in which robots can be involved in digital games: as a robot toy, as a board game play companion, and as an actor in pervasive games. Research work has already been carried out in the first two scenario types, though much of it is still some way from real-world application. The third scenario-type is still very much in the area of future research.

A common theme in all three is the difficulty in knowing the state of the human interaction partners involved. The level of difficulty is dependant on the exact application, with the iCat chess companion a good example of an application in which it is less difficult and a pervasive game one in which it is more difficult. Social signal recognition is related to basic configuration issues – whether the user is sitting, moving around, one or many; to the specific signals to be detected – easier for some facial expressions such as smiles, harder for many others; and above all to the social context. This context encompasses the overall social framework, which includes the rules and gameplay in the case of games, the specific interaction taking place in its game context, and the social roles involved.

A continuing research challenge is finding applications in which these factors allow successful incorporation of the game-playing robot without requiring functionality too far away from the current state-of-the-art.

As with other robotic applications, running an application over the long-term rather than in short lab-based experiments is also a challenge. This is partly related to obvious constraints like battery life and the need for recharging, as discussed earlier. Running complex robot software continuously over periods of weeks is a further technical challenge. However there are more interesting social issues. Robots have a pronounced novelty effect because most people rarely interact with them and have often never come across a real-world example. This produces an untypical social interaction over the short-term. Continuing social interaction beyond the novelty effect is a research issue that is only slowly being confronted.

Even a relatively successful game-playing robot such as the iCat chess companion was less successful as time passed and the novelty effect faded. Child players looked at it less, smiled at it less, and commented upon its behaviour less as time went on [23]. It is likely that this was related to the range and variety of its social behaviour and that over time it became too predictable. Thus developing the social repertoire of a game-playing robot requires careful thought if it is to succeed over the longer term.

This is a difficult design problem but also a difficult evaluation problem. Iterating and reiterating a design that may take several weeks to evaluate requires a substantial commitment of research resource as well as access to a large number of willing participants over long periods.

We finish by reiterating the generic requirements discussed above. Expressive behaviour underpinned by affective processing is at least as crucial to functional and engaging game-playing robots as an ability to play good moves.

References

1. Aylett RS (2004) Agents and affect: why embodied agents need affective systems. In: Methods and applications of artificial intelligence. Springer, pp 496–504
2. Aylett R, Kriegel M, Wallace I, Márquez Segura E, Mecurio J, Nylander S, Vargas P (2013) Do I remember you? Memory and identity in multiple embodiments. In: RO-MAN, 2013 IEEE. IEEE, pp 143–148
3. Breazeal C (2002) Designing social robots. MIT Press, Cambridge, MA
4. van Breemen A, Yan X, Meerbeek B (2005) iCat: an animated user-interface robot with personality. In: Proceedings of the fourth international joint conference on autonomous agents and multiagent systems, ACM, pp 143–144
5. Castellano G, Pereira A, Leite I, Paiva A, McOwan PW (2009) Detecting user engagement with a robot companion using task and social interaction-based features. In: Proceedings of the 2009 international conference on multimodal interfaces, ACM, pp 119–126
6. Dennett DC (1989) The intentional stance. MIT Press, Cambridge, MA
7. Dias J, Aylett R, Paiva A, Reis H (2013) The great deceivers: virtual agents and believable lies. Proc Cog Sci 2013:2189–2194
8. Donath J (2004) Artificial pets: simple behaviors elicit complex attachments. In: Bekoff M (ed) Encyclopedia of animal behavior
9. Duffy B (2000) The social robot. Ph.D. Thesis, Department of Computer Science, University College Dublin
10. EMOTE. http://www.emote-project.eu/. Accessed 27 Sept 2015
11. Fernaeus Y, Håkansson M, Jacobsson M, Ljungblad S (2010) How do you play with a robotic toy animal? A long-term study of pleo. In: Proceedings of the 9th international conference on interaction design and children, ACM, pp 39–48
12. Fernandez JMA, Bonarini A (2013) Towards an autonomous theatrical robot. In: Affective computing and intelligent interaction (ACII), 2013 Humaine Association Conference on (pp 689–694). IEEE
13. Fong T, Nourbakhsh IR, Dautenhahn K (2003) A survey of socially interactive robots. Robot Auton Syst 42(3–4):143–166
14. Fujita M (2004) On activating human communications with pet-type robot AIBO. Proc IEEE 92(11):1804–1813
15. Friedman B, Kahn PH Jr, Hagman J (2003) Hardware companions? What online AIBO discussion forums reveal about the human-robotic relationship. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 273–280
16. Kidd CD, Taggart W, Turkle S (2006) A sociable robot to encourage social interaction among the elderly. In: Robotics and automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference. IEEE, pp 3972–3976
17. Knol E, De Vries P (2011) EnerCities, a serious game to stimulate sustainability and energy conservation: preliminary results. eLearning Pap 25:1887–1542
18. Knight H (2011) Eight lessons learned about non-verbal interactions through robot theater. In: Social robotics. Springer, Berlin, pp 42–51
19. Kopp S, Krenn B, Marsella S, Marshall AN, Pelachaud C, Pirker H, Thórisson KR, Vilhjálmsson H (2006) Towards a common framework for multimodal generation: the behavior markup language. Proceedings of the IVA. Springer, pp 205–217
20. Kriegel M, Aylett RS, Cuba P, Vala M, Paiva A (2011) Robots meet IVAs: a mind-body interface for migrating artificial intelligent agents. IVA 2011 pp 282–295
21. LaFrance M (2008) What's in a robot's smile? The many meanings of positive facial display. Animating Expressive Characters for Social Interaction 74:37
22. Le QA, Pelachaud C (2012) Generating co-speech gestures for the humanoid robot NAO through BML. In: Gesture and sign language in human-computer interaction and embodied communication. Springer, pp 228–237

23. Leite I, Castellano G, Pereira A, Martinho C, Paiva A (2012) Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings. In: Proc HRI 2012. ACM, pp 367–374
24. Leite I, Pereira A, Martinho, C, Paiva A (2008) Are emotional robots more fun to play with? In: Robot and human interactive communication, 2008. RO-MAN 2008. 17th IEEE international symposium on. IEEE, pp 77–82
25. Lim MY, Aylett RS, Jones CM (2005) Emergent affective and personality model. IVA 2006. LNAI 3661:371–380
26. Lin CY, Cheng LC, Huang CC, Chuang LW, Teng WC, Kuo CH, Gu HY, Chung KL, Fahn CS (2013). Versatile humanoid robots for theatrical performances. Int J Adv Robotic Sy 10(7)
27. LIREC. http://lirec.eu. Accessed 22 Sept 2015
28. Magerkurth C, Cheok AD, Mandryk RL, Nilsen T (2005) Pervasive games: bringing computer entertainment back to the real world. Comput Entertain (CIE) 3(3):4–4
29. Mehrabian A (1977) Nonverbal communication. Transaction Publishers
30. de Melo C, Carnevale P, Gratch J (2011) Reverse appraisal: inferring from emotion displays who is the cooperator and the competitor in a social dilemma. In: Proc. Cog Sci 2011 pp 396–401
31. Ortony A, Clore G, Collins A (1988) The cognitive structure of emotions. Cambridge University Press, Cambridge
32. Pelachaud C (2005) Multimodal expressive embodied conversational agents, ICM 2005, pp 683–689
33. Pereira A, Prada R, Paiva A (2012) Socially present board game opponents. In: Advances in computer entertainment. Springer, Berlin, pp 101–116
34. Ribeiro T, Pereira A, Deshmukh A, Aylett R, Paiva A (2014) I'm the mayor: a robot tutor in enercities-2. In: Proc. AAMAS 2014. IFAAMAS, pp 1675–1676
35. Rodrigues SH, Mascarenhas S, Dias J, Paiva A (2014) A process model of empathy for virtual agents. Interact Comput, iwu001
36. Segura EM, Kriegel M, Aylett R, Deshmukh A, Cramer H (2012) How do you like me in this: user embodiment preferences for companion agents. Springer, Berlin, pp 112–125, IIVA 2012
37. Sharkey A, Sharkey N (2012) Granny and the robots: ethical issues in robot care for the elderly. Ethics Inf Technol 14(1):27–40
38. Snyder M, Swann WB (1978) Behavioral confirmation in social interaction: from social perception to social reality. J Exp Soc Psychol 14(2):148–162
39. Sundstrom P (2005) Exploring the affective loop. Stockholm University, Stockholm
40. Thiebaux M, Marsella S, Marshall AN, Kallmann M (2008) Smartbody: behavior realization for embodied conversational agents. In: Proc AAMAS-Vol 1. IFAAMAS, pp 151–158
41. Zeglin G, Walsman A, Herlant L, Zheng Z, Guo Y, Koval MC, . . . Srinivasa SS (2014) HERB's sure thing: a rapid drama system for rehearsing and performing live robot theater. In: Advanced robotics and its social impacts (ARSO), 2014 IEEE Workshop on (pp 129–136). IEEE
42. Zeng Z, Pantic M, Roisman G, Huang TS (2009) A survey of affect recognition methods: audio, visual, and spontaneous expressions. Pattern Anal Mach Intell IEEE Trans 31(1):39–58

Chapter 18
Lovotics: Love and Sex with Robots

Adrian David Cheok, David Levy, and Kasun Karunanayaka

Abstract The publication of the book "Love and Sex with Robots", late in 2007 by Dr. David Levy, heralded a new era in this somewhat controversial field. Human-Robot intimate relationships were no longer pure science fiction but had entered the hallowed halls of serious academic research. This book chapter presents a summary of significant activity in this field during the recent years, and predicts how the field is likely to develop. We will also detail about our research in physical devices for human-robot love and sex communication.

Introduction

The publicity department at Harper Collins in New York, who had published the original English language edition of the book "Love and Sex with Robots" [1], found an eager public in North America who wanted to know more. During the period immediately prior to publication of the book and for a few months afterwards, the topic caught the imagination of the media, not just in the USA and Canada but on a worldwide scale. During those months Dr. David Levy gave some 120 interviews, by telephone, email and in person; to newspapers, magazines, radio, and TV stations and to electronic media. Television interviews included an appearance on The Colbert Report [2]; as well as visits to Dr. David Levy's home by TV crews from Russia, Canada, Austria, France, Germany, Switzerland and other countries.

A.D. Cheok (✉)
City University London, London, UK

Imagineering Institute, Nusajaya, Malaysia
e-mail: adrian@imagineeringinstitute.org; adriancheok@mixedrealitylab.org

D. Levy
Imagineering Institute, Nusajaya, Malaysia

Retro Computers Ltd, London, UK
e-mail: david@retro-computers.co.uk

K. Karunanayaka
Imagineering Institute, Nusajaya, Malaysia
e-mail: kasun@imagineeringinstitute.org

There was also, not surprisingly, a flurry of interest from women's magazines, including Elle and Marie Claire. And the coverage in general science publications included articles in IEEE Technology and Society Magazine, MIT Technology Review, Scientific American, and Wired.

In the academic world there has already been sufficient coverage of the topic to demonstrate rather convincingly that it is of interest not only for mainstream media. An academically rewritten version of the book became Dr. David Levy's PhD thesis "Intimate Relationships with Artificial Partners" – at the University of Maastricht, where the thesis defense in October 2007 attracted more media publicity than any previous PhD in the university's history. Conferences on robotics, AI and other Computer Science related subjects began to accept and even invite papers on the subject, and there have thus far been two conferences devoted specifically to Human-Robot Personal Relationships. In 2014 the First International Congress on Love and Sex with Robots was held in Madeira. The academic journals that have since chosen to publish papers on the topic have included: Accountability in Research, AI and Society, Artificial Intelligence, Current Sociology, Ethics and Information Technology, Futures, Industrial Robot, International Journal of Advanced Robotic Systems, International Journal of Social Development, International Journal of Social Robotics, International Journal of Technoethics, New Media and Society, Phenomenonology and the Cognitive Sciences, Philosophy Technology, Social Robotics, Technological Forecasting and Social Change, and various publications from the IEEE, Springer and other highly respected technology stables. One paper, from Victoria University of Wellington, New Zealand, achieved a high profile in the general media when it appeared in 2012 for its entertaining depiction of a future scenario in the red light district of Amsterdam – a life, in 2050, revolving around android prostitutes "who are clean of sexually transmitted infections (STIs), not smuggled in from Eastern Europe and forced into slavery, the city council will have direct control over android sex workers controlling prices, hours of operations and sexual services" [3].

Since the initial burst of media interest late in 2007 there have also been TV documentaries and feature movies in which sex with robots, virtual characters, or with life-sized sex dolls was the dominant theme: Lars and the Real Girl, Meaning of Robots (which had its premiere at the 2012 Sundance Festival), My Sex Robot, Her (2013), and the BBC TV documentary Guys and Dolls as well as the 2004 remake of The Stepford Wives. This points out that it is the sexual nature of the subject matter which is responsible. Sex Sells.

Following the storm of publicity created by the launch of the Dr. David Levy's book and the defense of his thesis in 2007, the subject of human-robot romantic and intimate relationships rapidly developed into an academic research discipline in its own right. The subject was named "Lovotics", a term coined during discussions at the National University of Singapore between Adrian David Cheok, Sam Ge and Hooman Samani, and first mentioned in the literature in 2009 [4]. In his PhD thesis

in 2011, Adrian David Cheok's student Hooman Samani explored certain aspects of Lovotics, and describes the design and development of a hardware platform – a robot – which was capable of experiencing complex and human-like biological and emotional states that were governed by artificial hormones within its system [5]. Samani's robot was a novel advanced artificial intelligence system and is described in a little more detail in Sections below.

The interest in this field from the academic community resulted, in 2013, in the founding of a journal and e-journal devoted entirely to the subject, whose Editor-in-Chief is Adrian David Cheok. Lovotics [6] defines its own domain as "Academic Studies of Love and Friendship with Robots" (Fig. 18.1).
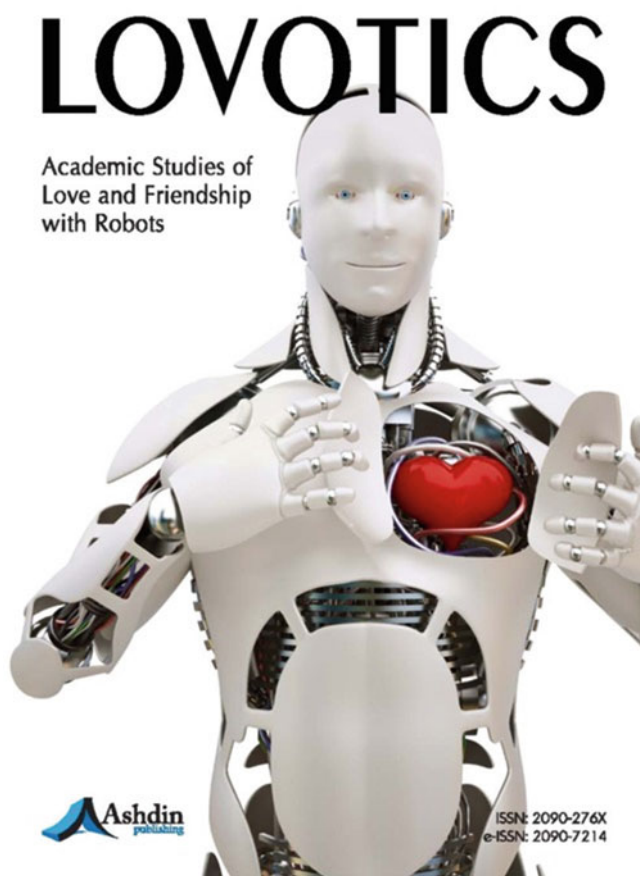
Fig. 18.1 Journal lovotics, academic studies of love and friendship with robots (Image used with permission)

The First Crude Sex Robot

One of the most often asked questions in media interviews with the Dr. David Levy in 2007–2008 was this: "How soon do you think the first sex robots will be on the market?" His consistent response was that the technologies necessary to create a crude sex robot were already available, and therefore it would probably not be more than 2–3 years before some enterprising entrepreneur(s) put these technologies together. For example, a sex doll with certain parts vibrating and some sexy synthetic speech would be a significant step up for those customers who have hitherto purchased just a static sex doll. This applies equally to a malebot as to a fembot – the worldwide commercial success of female vibrators indicates that a male sex doll endowed with a well-designed vibrating penis would be a good start in that direction.

Late in 2009 publicity began to appear in the media about a "sex robot" developed by a New Jersey entrepreneur, Douglas Hines. His web site www.truecompanion.com proudly proclaimed that

> We have been designing "Roxxxy TrueCompanion", your TrueCompanion.com sex robot, for many years, making sure that she: knows your name, your likes and dislikes, can carry on a discussion and expresses her love to you and be your loving friend. She can talk to you, listen to you and feel your touch. She can even have an orgasm!

Other amazing claims on the truecompanion.com site included:

> She also has a personality which is matched exactly as much as possible to your personality. So she likes what you like, dislikes what you dislike, etc. She also has moods during the day just like real people! She can be sleepy, conversational or she can "be in the mood"!

and

> Roxxxy also has a heartbeat and a circulatory system! The circulatory system helps heat the inside of her body.

and

> She can talk to you about soccer, about your stocks in the stock market, etc.

For millions of men eagerly awaiting the next major technological development that would enhance their sex lives, the announcements about Roxxxy probably seemed too good to be true. And they were! The press launch of Roxxxy took place at the Adult Entertainment Expo in Las Vegas on January 9th 2010, but it posed more questions than it answered. It appeared, for example, that touching Roxxxy's hand caused it to exclaim that "I like holding hands with you", but what does that prove? Only that an electronic sensor was linked to some sort of recorded sound output. It was not a demonstration of the speech technology that would be needed in a talking conversational robot. And furthermore, Hines's behaviour during the demonstration prompted the question – how much of the technology was inside Roxxxy and how much in the computer or whatever electronics were located behind the prototype?

The media hype surrounding Hines's launch in Las Vegas seems to have attracted the attention of many prospective customers for Roxxxy's supposedly seductive charms. At the beginning of February 2010 Hines's web site started to take orders for Roxxxy, advertising the product at a "sale price" of $6,495, which it claimed represented a reduction of $500. Accompanying the invitation to place an order, the site also presented a "Master Agreement" that extended to 15 clauses of legalese, covering the purchase of Roxxxy and subscriptions to associated services, but the "RETURNS, REFUNDS AND CANCELLATION POLICY" of that agreement (clause 12.1) made it clear that once production of a customer's Roxxxy commenced, the purchaser could not get any of their money refunded. This begs the question – why would any prospective customer be willing to part with their money without any possibility of recovery, when there had been no public demonstration or independent product review of a fully working Roxxxy that could perform as advertised?

Shortly after truecompanion.com started taking orders for Roxxxy, various news sites posted comments such as:

> Roxxxy won't be available for delivery for several months, but Hines is taking pre-orders through his Web site, TrueCompanion.com, where thousands of men have signed up.

Doubts about Roxxxy persist to this day (July 2015). Dr. David Levy wrote an exposé entitled "Roxxxy the 'Sex Robot' – Real or Fake?" and posted it on www.fembotcentral.com. And the Wikipedia entry for Roxxxy [7] includes the following:

> According to Douglas Hines, Roxxxy garnered about 4,000 pre-orders shortly after its AEE[1] reveal in 2010. However, to date, no actual customers have ever surfaced with a Roxxxy doll, and the public has remained skeptical that any commercial Roxxxy dolls have ever been produced.

If it is true that Hines received 4,000 pre-orders then he would have raked in something over $20 million for those orders, since his web site demands payment in advance. But as the above extract from the Wikipedia entry indicates, neither Hines himself or any of his customers has demonstrated, in public or to reputable media, the advertised features of Roxxxy actually working. Three years after its "launch" there still appears to be absolutely no sign of a demonstrable product that can talk about Manchester United (as Hines claimed Roxxxy could do) or perform in the other ways that Hines's advertising blurb claimed for Roxxxy.

Despite all the negative aspects of Hines's operation and of the product itself, the launch of Roxxxy at the January 2010 Adult Entertainment Expo can be viewed as some sort of milestone – a vindication of the forecast for a 2–3 year time span from late 2007 to the launch of the world's first commercially available sex robot. Hines has proved that there is indeed a significant level of interest in sex robots from the buying public.

---

[1] Adult Entertainment Expo

Lovotics

Hooman Samani's PhD thesis describes the design and development, of a robot aimed at imitating the human affection process so as to engender attraction, affection and attachment from human users towards the robot [5]. In his thesis abstract Samani summarizes the design of the robot thus:

> The artificial intelligence of the robot employs probabilistic mathematical models for the formulation of love. An artificial endocrine system is implemented in the robot by imitating human endocrine functionalities. Thus, the robot has the capability of experiencing complex and human-like biological and emotional states as governed by the artificial hormones within its system. The robot goes through various affective states during the interaction with the user. It also builds a database of interacting users and keeps the record of the previous interactions and degree of love.

The artificial intelligence of the Lovotics robot includes three modules: the Artificial Endocrine System, which is based on the physiology of love; the Probabilistic Love Assembly, which is based on the psychology of falling in love; and the Affective State Transition, which is based on human emotions. These three modules collaborate to generate realistic emotion-driven behaviors by the robot.

The next four subsections summarize the formulation of love that underpins much of Samani's work, as well as the three software modules of the system mentioned above. The combined effect of these modules is to provide an artificially intelligent model that can display a range of emotions, adjusting its affective state according to the nature and intensity of its interactions with humans. The goal is to develop a robotic system that can exude affection for the user and react appropriately to affection from the user.

The Formulation of Love

The robot's intimacy software employs parameters derived and quantified from five of the most important reasons for falling in love [1]: proximity, repeated exposure, attachment, similarity and attraction. Intimacy in the robot is thereby related to those same factors that cause humans to fall in love. The robot utilizes audio and haptic channels in order to provide these different types of input which communicate the user's emotional state to the robot [8]. The audio channel carries data for five audio parameters that characterize emotional cues within a human voice. The haptic channel carries data relating to the user touching the robot – the area of contact between robot and human and the force of that touch.

The Lovotics robot includes mathematical models for those five causal factors of love, creating a mathematical formula to represent each factor as well as a single "overall intimacy" formula which combines these five individual formulae into one. As an example of the five models, the proximity formula incorporates various

distances between robot and human that indicate, *inter alia*, how closely the robot and human are to touching each other, and how close they are emotionally.

The Probability of Love

The robot algorithm has taken account of the various factors that can engender human love, in order to develop a systematic method for assessing the level of love between a robot and a human. This is achieved by formulating probabilistic mathematical models for these factors, which in turn enable the robot to determine the level of intimacy between humans and robots. These models can be represented in a Baysian network that depicts the relationship between love and its causal factors. The factors involved in this model include: proximity – the physical distance between human and robot; propinquity – spending time with each other; repeated exposure – this can increase familiarity and liking in the other individual; similarity – this is directly related to the feeling of love; etc.

The probabilistic nature of these parameters allows a Baysian network to be employed to link the parameters to relevant audio, haptic and location data, leading to an estimate of the probability of love existing between robot and human. For example, audio proximity is employed in the calculations to emulate the effects of physical distance.

From the various causal parameters, the system calculates the probabilistic parameters of love, resulting in an appraisal of the level of love between human and robot [9].

The Artificial Endocrine System

The human endocrine system is a system of glands that secretes different types of hormones into the bloodstream. The purpose of those hormones is to maintain homeostasis, i.e. to regulate the internal environment of the body in order to keep certain functions stable, such as body temperature, metabolism and reproductive functions.

The Lovotics artificial endocrine system is based on the human endocrine system, employing artificial hormones to create a simulation of the human system. The artificial hormones are software simulations of those human hormones which are related to the emotions – Dopamine, Seratonin, Endorphin and Oxytocin, inter alia. The levels of these artificial hormones change dynamically due to the robot's interactions with users and according to its awareness of its emotional and physical circumstances.

The Affective State Transmission System

The affective state of the Lovotics robot depends largely on the various inputs it receives that are caused by its interactions with humans. Every interaction provide input data that is mapped onto a combination of six basic emotional parameters: happiness, sadness, disgust, surprise, anger and fear. These six emotions are widely employed and described in the emotion literature.

The manner in which the robot's emotional state changes with the various inputs it receives is controlled by a model of emotion referred to as Affective State Transition [9]. The Lovotics robot has a novel transition system which governs the immediate emotional changes in the robot. Their transition system functions in collaboration with the "probabilistic love assembly" module in order to control the overall emotional state of the robot. The short-term affective state of the robot is thereby transformed repeatedly into other affective states which are determined by the robot's previous affective states, its current mood, and the influences of the various input data it received during its interactions with humans, including audio and touch, and with its environment. For example, temperature could be one environmental input that might be programmed to influence the robot's affective state, if it "dislikes" being cold.

The Kissenger

In order for robots, such as the Lovotics robot, to have realistic physical interactions with humans, technology needs to be developed for human – machine kissing. Thus, we have developed a kissing robot messenger or "Kissenger" (Fig. 18.2).

We live in a global era, and more and more couples and families are apart due to work and business. New technologies are often employed to help us feel connected with those whom we care about, through an increasing interest in touch and feeling communication between humans in the human-computer interaction community. Research like "Hugvie" [10] and the "Hug over a Distance" project [11] tested the feasibilities of telepresence and intimacy technology. However, these are big, bulky, and impractical.

There is some commercial work like "Hug Shirt" [12] and "Huggy Pajama" [13], which explore hugging in remote with loved ones using wearable fashion technology. But these still lack a proper interface for "abstracted presence". Thus, we propose a new system to feel the real presence using communication over internet for humans or robots.

Kissing is one of the most important modes of human communication as it conveys intimacy and many deeply feel positive emotions such as respect, greeting, farewell, good luck, romantic affection, and/or sexual desire through the physical joining or touching of lips by one individual on another individual's cheek, forehead, lips, etc. [14].
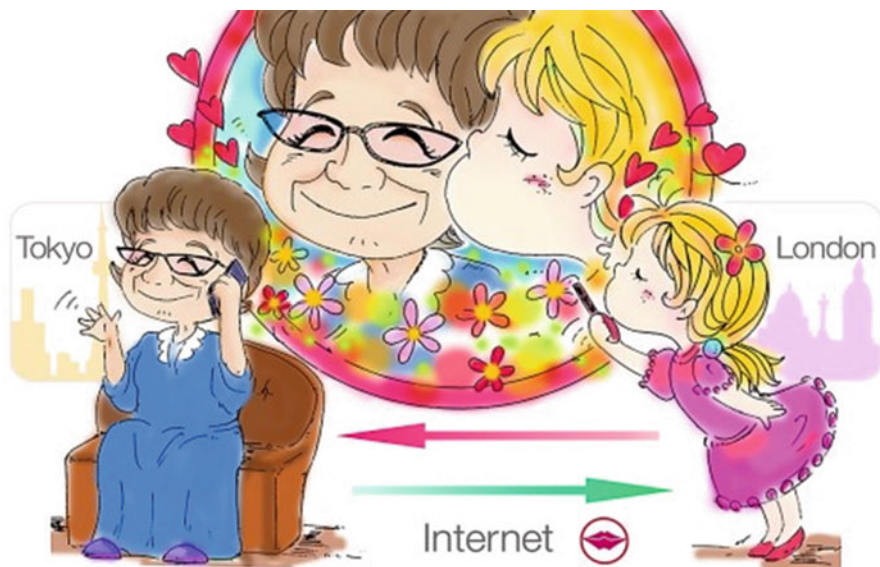
Fig. 18.2 The concept of kiss communication

The first practical device to be developed by the Lovotics community in Asia was unveiled at the Designing Interactive Systems Conference in Newcastle in June 2012 [15]. Its development was started in the Mixed Reality Lab by Professor Adrian David Cheok and his team. It would be possible to integrate the Kissenger technology into a sex robot but initially its use will be in teledildonic products for enabling lovers to kiss each other via the Internet.

The Kissenger employed soft, pressure sensitive, vibrating silicone lips which, in the early prototypes, stood out from the surface of a smooth plastic casing shaped somewhat like a human head. Those early prototypes have since been replaced by a version for mobile phones.

Considering this missing dimension in today's communication technologies, we aim to design a new device to facilitate the exchange of emotional content to feel a closer sense of presence between people who are physically separated, thus integrating their interpersonal relationships further.

When a user kisses the device on its lips, the changes in shape of the lips are detected by sensors and the resulting data is transmitted over the Internet to a receiving Kissenger, which converts the data back to lip shapes. This reproduces the changes in the kisser's lip shape, changes which are felt by the kisser's partner.

The Kissenger technology could perhaps be enhanced with an idea from a rather more ambitious haptic device of the same ilk which has been developed in Tokyo at the Kajimoto Laboratory in the University of Electro-Technology. Their invention is a French-kissing device whose prototypes are not yet at a stage where they are likely to inspire erotic thoughts, being based on a straw-like tube that moves when
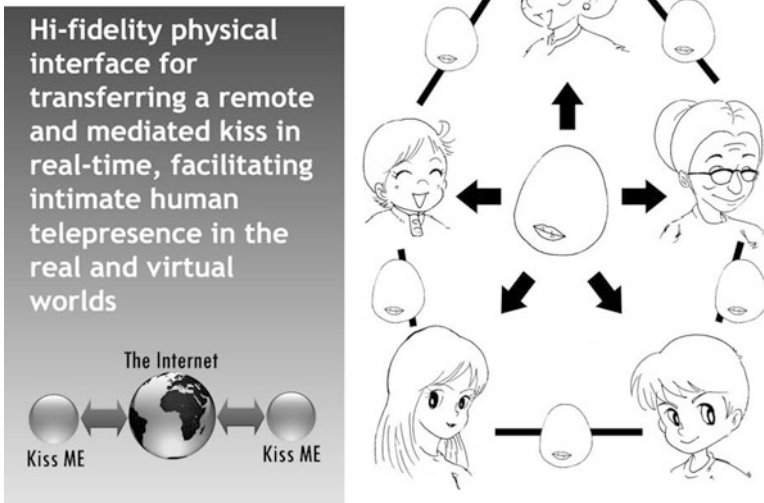
Fig. 18.3 Kissenger usage scenario A

in contact with a user's tongue. But we can expect to see an enhanced form of this idea in a future version of the Kissenger and similar inventions – enhancements under consideration at the Kajimoto Laboratory include adding taste, breath and moisture to the experience.

During a kiss, along with its strong emotional and affectionate connections, a series of physical interactions takes place. The touch of the lips exchanges the pressure, softness, and warmth of each lip in a convincing way. We approached this design problem carefully, given the intimate nature of the interaction and iteratively designed Kissenger which consists of two paired devices that can send and receive kisses simultaneously as shown in concept images Figs. 18.3 and 18.4.

After studying the biological and psychological parameters of a kiss, a series of exploratory form factors were drawn to help visualize the possible interfaces. Figure 18.5 shows some of our initial concept designs.

At this stage, we looked for designing a system that effectively transmits the same sensation of kiss to one another. The one key issue was that the use of the device should be comfortable and not distract or obstruct the natural interaction of the kiss. Hence, we decided to integrate the initial concept design for a lip-like portable device with a minimalistic shape. However, one of the main concerns was the lip needed to be equipped with sensors and actuators. Hence, we looked into the possible technologies and sizes which could be fit into the form factor of our device. Figure 18.6 shows the 3D depiction of the proposed device with a new shape which can attach to a smart phone, allowing a video call and virtual kiss simultaneously.
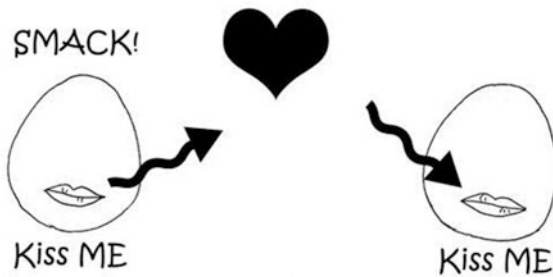
Fig. 18.4 Kissenger usage scenario B

Design Features

The interaction mechanism for Kissenger was devised with a number of features that we believe will make kiss communication between two users more meaningful. The system consists of following key features:

- Lip sensor push and pull reverse feedback for kiss behavior.
- Lip rotation force feedback
- Sending scents.
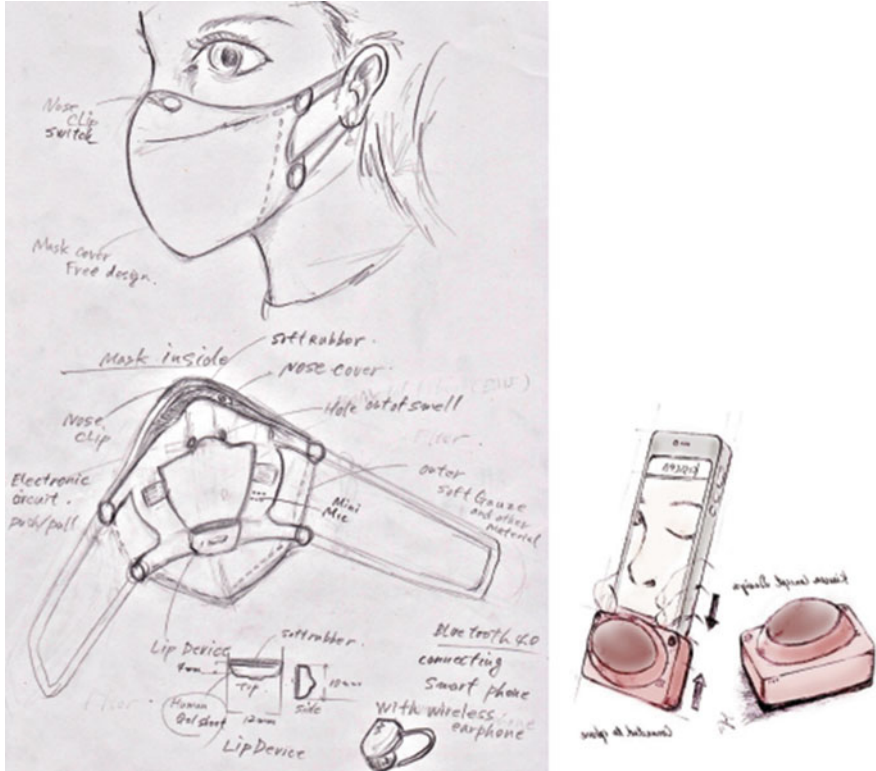
Fig. 18.5 Preliminary concept designs of Kissenger
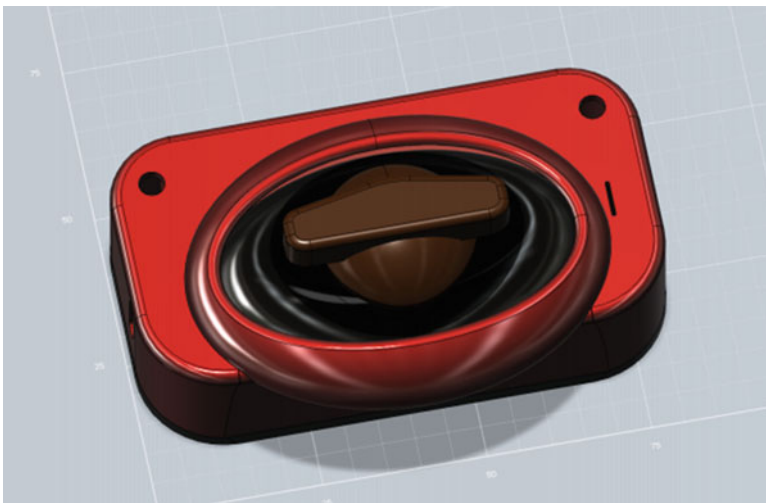
Fig. 18.6 The new design of Kissenger which can attach to a mobile phone

- Feeling LED light color communication (red, orange, green, and blue)
- Apps for kiss communication with video chat (Facetime, Google + Hangouts, Skype, Facebook, Zoom, etc. . . .)
- Changing the user characters and voices (face images)
- One-to-one pair and one-to-many user connections
- Recording the behavior of the partner's lips.
- Scent tank changes the scent to suit the partners.
- Soft silicone cover made with gel for kiss communication.

Design Flow

The hardware design of Kissenger with all the features listed above specifies the use of a LED light, pressure sensors, actuators, a vibration motor, a scent tank, and a smartphone connector in the Kissenger design flow. Their design role is as follows:

Input Kiss Sensing

The front of the lip has pressure sensors placed just below the outer surface to initiate the Kissenger for the transmitter (the kissing person), the receiver (the kissed person), and also to sense varying levels of soft touches. The features for the lip sensor push and pull reverse feedback for kiss behavior as shown in Fig. 18.7.

Upon initialization, the front end of Kissenger can be tilted to a maximum of 18° to replicate different styles of kissing. Thus, this design simplifies the interface and enables users to form a correct and semantically meaningful mental representation of the system with great feasibility for real kissing. The system also can be used for kissing a robot or a virtual 3D character.

User A   User B
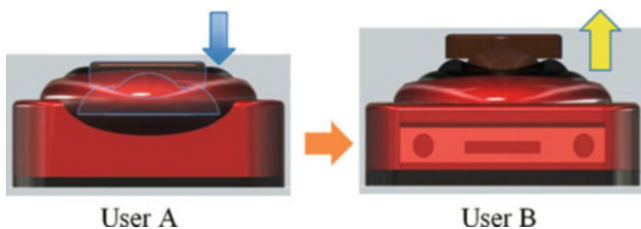
Fig. 18.7 Lip sensor push and pull reverse feedback for kiss behavior

Control and Wireless

Each Kissenger device is equipped with lip sensors (pressure sensor + heat sensor), a scent tank, a smartphone connector and voice speaker (Fig. 18.8) connected to an embedded circuit that controls the sensors and actuators with your phone and thereon with other Kissenger devices through the internet. Data from the pressure sensors is read continuously until a change is detected. If there a substantial change, the resulting increase is transmitted wirelessly to a receiver circuit that then actuates a servo motor array to produce similar motion of the lips.

Output Kiss Actuation

The kiss sensation on receiver (the kissed person) is produced through the movement of servomotors that distend the surface of the lip. Simultaneously, the scent tank. LED light, and voice speaker are actuated with pheromones, colors and sounds respectively to depict different moods (Fig. 18.9). Pheromones are the scents used in Kissenger that are capable of acting outside the body of the secreting individual to impact the behavior of the receiving individual, giving the feel of real presence of the partner. The shape and size of the lip cover hide the inner electronics that go into the sensing, control, and actuation of the device. Thus all these features make the user more amicable to this device and help evoke emotional responses and feelings for kiss communication.

Communication

Two or more Kissenger devices are wirelessly connected to each other via the smartphone Kissenger app, which are internally connected to their respective Smartphones as shown in Fig. 18.10. One of the unique added features of the app is that it allows one-to-many user communication along with one-to-one user communication as shown in Fig. 18.11. With the Kissenger app, the user can also actuate and transmit different colors to their partners to depict different moods with different scents thus giving a real sense of kissing.

An assessment of the new proposed shape and its implementation was conducted with a wide variety of people including researchers not involved in our project, mall shoppers, and friends over a period of time with around 50 people from different cultural backgrounds, age, and sexes participated in the evaluation process and provided feedback for the proposed shape and features. The major feedback is to integrate the size to make it more portable and user-friendly and to provide room for asynchronous kissing. There is the ability for the device to store a kiss that can be read at a later time on which we will be working in the future for the social impact of our project.
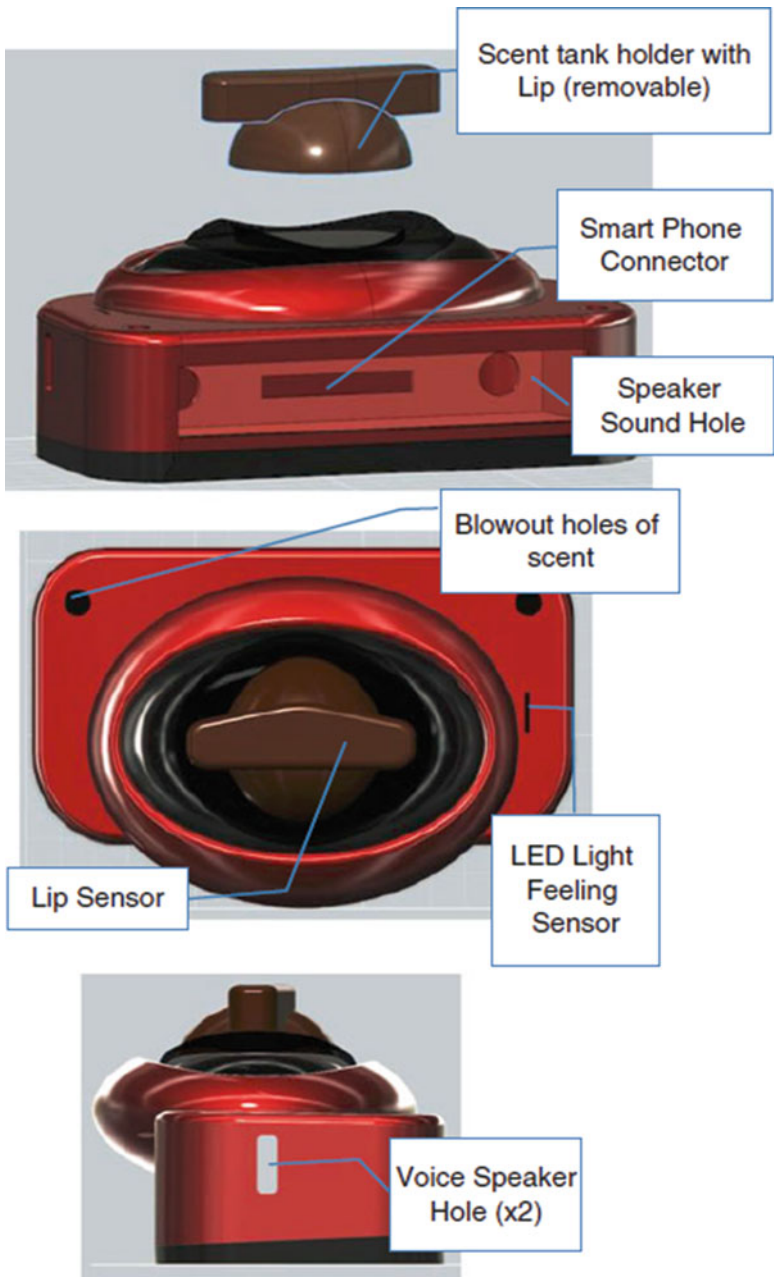
Fig. 18.8   Key design features of Kissenger

LED Color	User's Feeling Image
	Blue: (Chat start) Feeling: Normal Chat Time: 0s~120s Number of Kiss 0 Number of Sent 0
	Green: Feeling: Peace Chat Time: 120s~240s Number of Kiss 1~3 Number of scent 1
	Orange: Feeling: Empathy Chat Time: 240s~300s Number of Kiss 3~5 Number of scent 2
	Red: Feeling: Love (sincerely) Chat Time: 300s~ Number of Kiss 5~ Number of scent 3 more (1/s)

Fig. 18.9 LED light feeling sensor color depiction
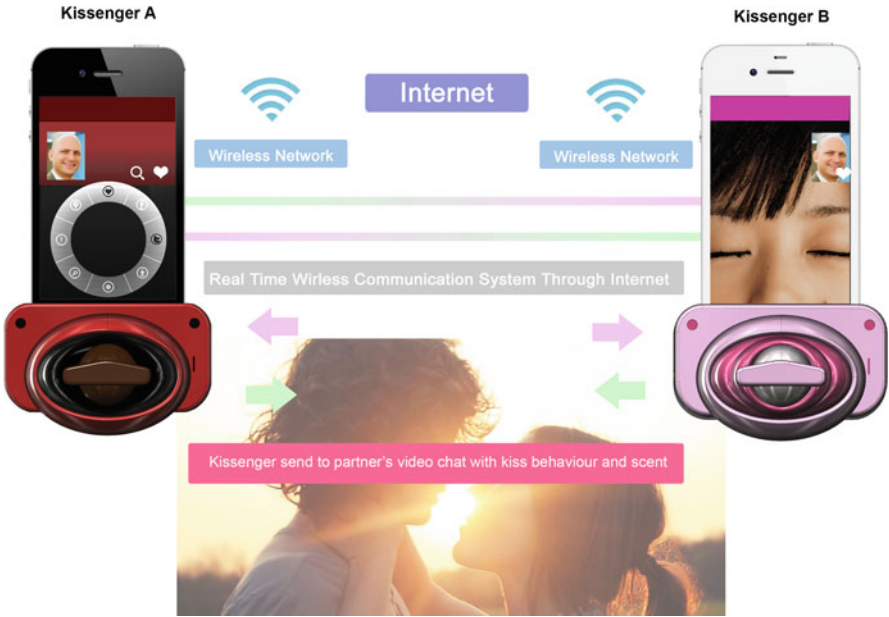
Fig. 18.10 Kissenger system diagram
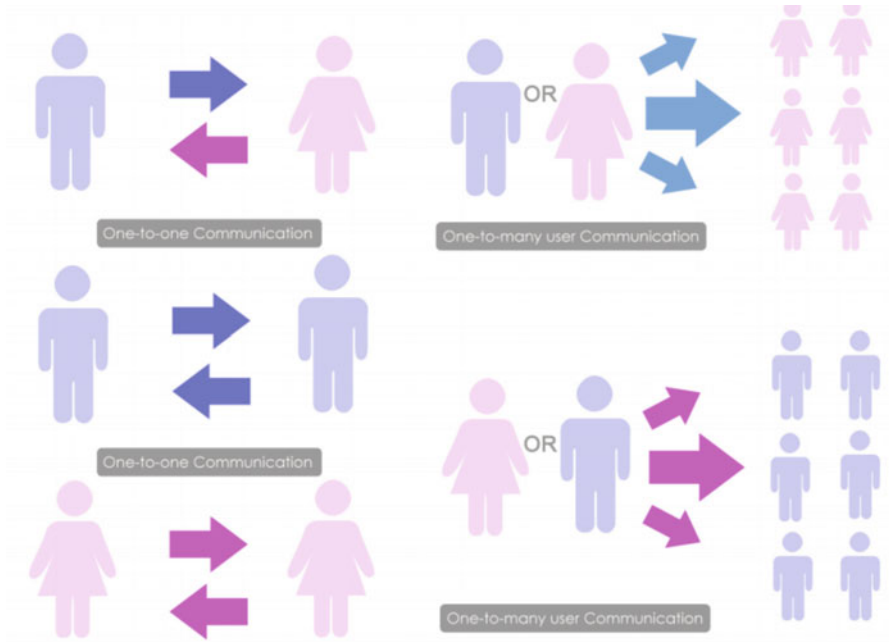
Fig. 18.11 User communication via Kissenger app

The Ethical and Legal Debate

The ethics of robot sex were first aired in an academic forum at the EURON Workshop on Roboethics in 2006 [16–18]. The following year, Dr. David Levy has discussed five aspects of the ethics of robot prostitution at an IEEE conference in Rome [19]: the ethics of making robot prostitutes available for general use; the ethics vis à vis oneself and society in general, of using robot prostitutes; the ethics vis à vis one's partner or spouse, of using robot prostitutes; the ethics vis à vis human sex workers, of using robot prostitutes; and the ethics vis à vis the sexbots themselves, of using robot prostitutes. Since the last of these issues is only of significance if robots are eventually developed with (artificial) consciousness, it is also relevant when considering this particular issue to contemplate the ethical treatment in general of artificially conscious robots [20].

A somewhat broader airing of the ethical impacts of love and sex machines was presented by John Sullins in 2012 [21]. Sullins explores the subject partly on the basis that such entities are programmed to manipulate human emotions "in order to evoke loving or amorous reactions from their human users". He submits that there should be "certain ethical limits on the manipulation of human psychology when it comes to building sex robots", and accordingly he identifies three design considerations which he proposes should be applied to the development of robots designed for love: (i) robots should not fool people into ascribing more feelings to the machine than they should; (ii) robot designers should be circumspect in how their inventions exploit human psychology; and (iii) robots should not be designed that intentionally lie to their users in order to manipulate their user's behaviour.

A considerably more strident attitude to the ethics of robot sex pervades a 2012 paper by Yusuff Amuda and Ismaila Tijani [22], which views the subject from an Islamic perspective. These authors appear to have no doubts that "having intercourse with robot is unethical, immoral, uncultured, slap to the marriage institution and respect for human being." While many might not concur with the robustness of their position on the subject, it cannot be denied that the question of robot sex within the confines of marriage, or indeed within any existing human sexual relationship, is a serious issue. The question most often asked of the present author in media interviews has been: "Is it cheating for someone who is married or in a committed relationship to have sex with a robot?"

In this author's opinion the answer is a resounding "No". A partner or spouse who has sex with a robot is no more guilty of cheating on their other half than are any of the tens of millions of women who use a vibrator. But not everyone agrees with this position, and in parallel with the possibility that sex with a robot should be regarded as cheating on one's spouse, there comes an interesting legal question which has been flagged by the California lawyer Sonja Ziaja [23]. Could a sex robot be legally regarded as the enticing protagonist in a law suit brought for the enticement of one's spouse? In the eight states of the USA where this type of law is still on the statute books, where they are called amatory or "heart balm" laws,

Ziaja questions whether a sex robot could be held to be the cause, or a contributing cause, to the breakdown and dissolution of a marriage, and if so, who should be held legally liable to pay whatever damages a court might assess? Ziaja suggests a few obvious possible culprits for cases of enticement by a robot: the robot's inventor, its manufacturer, its owner, or even the robot itself. But the attribution of liability for a wrong wrought by a robot is an extremely complex issue, one which this author believes will not be adequately solved in the foreseeable future. Instead it has been suggested [24] that robot wrongs could be compensated by an insurance scheme, much akin to that which works well for automobiles and other vehicles.

The only form of punishment considered by Ziaja for transgressing the American heart balm laws is to compensate the plaintiff, which is a notion that pales into insignificance when compared to the punishments discussed by Amuda and Tijani. They point out that, under Sharia law, judges are permitted to invoke lashes or even capital punishment for having sex with a robot, provided there is sufficient credible evidence of the crime [23]. "To this study, death penalty by hanging may not be applicable and implemented unless there are enough and credible evidences to justify the death by hanging of robot fornicator or adulterer."

Ziaja's paper largely avoids discussing punishment in relation to enticement cases in which a robot is the protagonist, preferring to prevent the problem from occurring by having robots designed in such a way as to incorporate feelings of heartbreak together with the goal of caring for those in its owner's circle of friends and relatives. "In order for robots to enter into human romantic relationships in a way that is consistent with the values underlying the heart balm torts, it may also need to experience heartache and empathy as we do." Ziaja's position thus supports that of John Sullins.

An in-depth consideration of whether or not human-humanoid sexual interactions should be legally regulated was discussed by Anna Russell in Computer Law and Security Review [25]. The very fact that such a discussion should appear in the pages of a respected legal journal points to the seriousness with which the legal profession is viewing the legal implications of the human-robot relationships of the future. Russell suggests that:

> "Regulation of human-humanoid sexual interaction either by the state or federal government[2] will be sought when the level of interaction either (1) mimics human sexual interactions currently regulated or (2) will create a social harm if the interaction is not regulated ... currently, in places where humans are using robots for pleasure in a sexual way that pleasure is either not regulated or is regulated in the way the use of any sexual device may be regulated" but that when more advanced robots – humanoids – are used for sexual pleasure, "then in many places, traditional norms and social mores will be challenged, prompting the development of state regulation. Will such regulation, then, be at odds with accepted notions of rights and freedoms?"

---

[2]In the USA.

Russell then delves further into the question of how regulation of human-humanoid sexual encounters would work, and highlights some of the questions that will arise, including:

> How many rights will humans allow if humanoids clamor for sexual freedoms? How will humanoids be punished for sexual transgressions? Will humanoids need legal protection from the abuse of human sexual proclivities?

Russell's conclusion is a call for the

> ... early discussion of the ramifications of a future species' demand for legal rights ... the legal profession should develop legal arguments before attest case occurs in order to avoid the illogic and danger of arguments that stem from species bias.

In 2011 the MIT Technology Review conducted a poll on people's attitudes to the idea of loving a robot. Nineteen percent of those questioned indicated that they believed they could love a robot, 45 % said "No" and 36 % responded "Maybe". When it came to a question of whether or not people believed that robots could love humans, 36 % said "Yes", only 23 % responded "No", and 41 % "Maybe". So already the idea of human-robot love was taking root as a serious proposition.

In a later poll, this one about robot sex rather than robot love, which was conducted in February 2013 by The Huffington Post and YouGov among 1,000 American adults, 9 % of respondents indicated that they would have sex with a robot, and 42 % opined that robot sex would constitute cheating on one's human partner (31 % said "No" to the cheating question, while 26 % said they were uncertain). This can be taken as further evidence that a significant portion of the population already regards robot sex as a serious subject. Just how serious can perhaps be judged by a news story that hit the media in March 2013 about an online auction for the virginity of a Brazilian sex doll called Valentina [26] which was inspired by a 20-year-old Brazilian woman, Catarina Migliorini, who had auctioned her own virginity for $780,000 (sold to a Japanese buyer). True, a sex doll is only an inanimate product, lacking all the interactive capabilities of the sex robots of the future. But the level of interest demonstrated by this news story bodes well for the commercial possibilities of sex robots.

For the Brazilian sex doll auction, the online retailer Sexônico offered a complete "romantic" package for the successful bidder, which included a one-night stay with Valentina in the Presidential Suite at the Swing Motel in Sao Paulo, a candlelit champagne dinner, an aromatic bath with rose petals, and a digital camera to capture the action. If the successful bidder lived outside Sao Paulo, Sexônico also offered to provide a round trip air ticket. Valentina's charms were not able to match the great commercial success of Ms Migliorini, but considering that most sex dolls retail at prices in the range $5,000–10,000 the final bid of $105,000 was still a good result for Sexônico, not to mention the value of all the media exposure they attracted.

Robot Love

In parallel with the developments we have discussed in the field of robot sex and teledildonics, there is a continuing and burgeoning research interest in robot love. Amongst the fundamental conditions for engendering human love, physical appearance and attractiveness rank highly. The translation of these conditions to the field of robotics have a champion in Professor Hiroshi Ishiguro, whose research teams are based at the Graduate School of Engineering Science at Osaka University and at the Hiroshi Ishiguro Laboratory in the Advanced Telecommunications Research Institute International in Kyoto.

Ishiguro is famous for, inter alia, the amazingly lifelike robots he has developed in various human images [27]. These include one in his own image which is sometimes sent to deliver his lectures when he is too busy to do so himself. Another of his robots, called "Geminoid-F" (Fig. 18.12), is made in the image of an attractive young woman who can blink, respond to eye contact, and recognize and respond to body language [28]. Ishiguro is encouraged in this aspect of his work by his conviction that Japanese men are more prone than are western men to develop amorous feelings towards such robots because, in Japan, with the influence of the Shinto religion, "we believe that everything has a soul and therefore we don't hesitate to create human-like robots".

Another strand of Ishiguro's research into artificially engendering feelings of love in humans is concerned with promoting romantic forms of communication. The "Hugvie" [29] is a huggable pillow, shaped in a somewhat human form, that is held by a user close to their body while they speak to their human partners

Fig. 18.12 "Geminoid-F" robot

via their mobile phone, located in a pocket in the Hugvie's head.[3] The Hugvie incorporates a vibrator to simulate a heartbeat, and the vibrations emanating from it are synchronized with the sounds of the partner's voice. This allows the simulated heartbeat to be changed according to the volume of the partner's voice, with the result that the listening user feels as though they are close to their partner. The comfort felt by holding the cushion, the sense of hugging one's partner, hearing one's partner's voice close to one's ear, and the simulated heartbeat aligned with that voice, all these combine to create a sense that the partner is in some way present, which in turn intensifies the listener's feelings of emotional attraction for their partner. Ishiguro expects this intensified affinity to increase the sense of intimacy between couples who are communicating through their respective Hugvies. Ishiguro shared in a breakthrough study that the Hugvie could decrease blood cortisol levels, therefore reducing stress [30]. Integrating the Hugvie technology into the design of an amorous robot might therefore enable a human user of such a robot to experience an enhanced feeling of a humanlike presence and a greater sense of intimacy from and for the robot.

Yet another direction of Ishiguro's research into having a robot engender emotions in humans is his investigation of the emotional effects, on a human user, of different facial expressions exhibited by a robot [31]. That research is currently in its early stages but there is already some indication that it will be possible for robots, by their own facial expressions, to affect a user's emotional state. Emotional facial expression is also a hot topic at the MIT Media Lab, where the Nexi robot was developed [32].

Predictions

Robot Sex

Clearly a significant sector of the public is now ready for the advent of commercially available sex robots, and the public's interest in and appetite for such products seems to be growing steadily. We have noticed a steady increase in the number of requests for media interviews on the subject during the past 2 years. Also growing steadily is the interest within the academic research community.

In our opinion nothing has occurred since the publication of Love and Sex with Robots to cast doubt on his 2007 prediction that sophisticated sex robots would be commercially available by the middle of this century. On the contrary, the increase in academic interest in this field has reinforced Dr. David Levy's conviction regarding that time frame.

What will be the next significant steps in this field? Intelligent electronic sex toys are gaining in popularity, for example the Sasi Vibrator, which "comes pre-loaded

---

[3]The Hugvie project grew out of an earlier Ishiguro project called "Telenoid"

with sensual intelligence which learns movements you like, specifically tailoring a unique experience by remembering movements that suit you."; and the "Love Glider Penetration Machine" which can be purchased from Amazon.com at around $700 and which is claimed to "give you the most comfortable stimulating ride you will ever have!" The Amazon web site also offers a very much more primitive looking sex machine at around $800, a machine of the type seen in many variations on the specialist site www.fuckingmachines.com, and which "supports multiple positions, has adjustable speeds, strong power, remote control.[4]"

Another research direction that perhaps offers even greater commercial potential comes from a combination of augmented reality with digital surrogates ("dirrogates") of porn stars. A recent (June 2013) posting by Clyde DeSouza [33] posits that the 3D printing of human body parts will enable the downloading, from "harddrives in Hollywood studios" of "full body digital model and 'performance capture' files of actors and actresses". DeSouza continues:

> With 3D printing of human body parts now possible and blue prints coming online with full mechanical assembly instructions, the other kind of sexbot is possible. It won't be long before the 3D laser-scanned blueprint of a porn star sexbot will be available for licensing and home printing, at which point, the average person will willingly transition to transhuman status once the 'buy now' button has been clicked.
>
>   . . .
>   If we look at Digital Surrogate Sexbot technology, which is a progression of interactive porn, we can see the technology to create such Dirrogate sexbots exists today, and better iterations will come about in the next couple of years. Augmented Reality hardware when married to wearable technology such as 'fundawear' [34] and a photo-realistic Dirrogate driven by perf-captured libraries of porn stars under software (AI) control, can bring endless sessions of sexual pleasure to males and females.

Fundawear is a prime example of the increase in popularity of intelligent electronic sex toys and teledildonic devices. It is a wearable technology project currently under development by the condom manufacturer Durex, which allows lovers to stimulate their partner's underwear via their respective mobile phones. Such products seem likely to benefit from the increased academic interest in Lovotics, which will surely lead to at least some of the academic research in this field being spun off into commercial development and manufacturing ventures. And the more prolific such products become in the market place, the more the interest in them and in fully fledged sex robots will grow. How long will it be before we see a commercially available sexbot much more sophisticated than Roxxxy? Almost certainly within the next 5 years.

---

[4]The sole review on Amazon.com as of May 2013 suggests that this product is poorly made and describes it as "a piece of junk".

Robot Love

The past few years have seen a surge of interest in research projects aimed at different aspects of love-with-robots. One aspect is concerned with enabling humans to convey amorous feelings to artificial partners, or to remotely located human partners with whom they communicate by artificial means (i.e. technology). Another aspect works in the opposite direction, enabling artificial partners to exhibit their artificial feelings, including love, to human partners. Some of this research has already demonstrated promising results, for example the experiments conducted with Hugvie by Ishiguro and his team in Japan. They plan further research with the Hugvie to investigate how vibration can further enhance the feeling of presence experienced by a user. Additionally they plan to employ tactile sensors to monitor the emotional state of a user, which will provide feedback for the Hugvie and thereby enhance its ability to influence a user's emotions. Ishiguro's team has already found that hugging and holding such robots "is an effective way for strongly feeling the existence of a partner".

Another domain to become an important catalyst for the development of human-robot emotional relationships, is what might be called girlfriend/boyfriend games. An example of this type of game is "Love Plus", which was first released in 2009 for the Nintendo DS games console, and subsequently upgraded for re-release. A recent (February 2013) article describes the relationship between a 35-year-old Tokyo engineer, Osamu Kozaki, and his girlfriend Rinko Kobayakawa [35]. When she sends him a message

> ... his day brightens up. The relationship started more than three years ago, when Kobayakawa was a prickly 16-year-old working in her school library, a quiet girl who shut out the world with a pair of earphones that blasted punk music.

Kozaki describes his girlfriend's personality as being

> ... the kind of girl who starts out hostile but whose heart gradually grows warmer. And that's what has happened; over time, Kobayakawa has changed. These days, she spends much of her day sending affectionate missives to her boyfriend, inviting him on dates, or seeking his opinion when she wants to buy a new dress or try a new hairstyle.

But while Kozaki has aged, Kobayakawa has not. After 3 years, she's still 16. She always will be. That's because she is a simulation; Kobayakawa only exists inside a computer.

Kozaki's girlfriend has never been born. She will never die. Technically, she has never lived. She may be deleted, but Kozaki would never let that happen.

Because he's "in love."

Conclusion

In this chapter, we discussed about the possibility of human robot intimate relationships and humanoid robot sex. We detailed Lovotics, which is a new research field that studies emotions of robots with an artificial endocrine system capable

of simulating love. We also presented the design and principle of Kissenger, an interactive device that provides a physical interface for transmitting a kiss between two remotely connected people. Finally we have discussed ethical and legal background and future predictions of love and sex with robots.

References

1. Levy D (2007) Love and sex with robots. Harper Collins, New York
2. The Colbert Report (2008) http://www.colbertnation.com/the-colbert-report-videos/147893/january-17-2008/david-levy
3. Yeoman I, Mars M (2011) Robots, men and sex tourism. Futures 44(4):365–371
4. Nomura S, Teh K, Samani H, Godage I, Narangoda M, Cheok A (2009) Feasibility of social interfaces based on tactile senses for caring communication. In: The 8th international workshop on Social Intelligence Design – SID 2009
5. Samani H (2011) Lovotics: love + robotics, sentimental robot with affective artificial intelligence. PhD thesis, National University of Singapore
6. Lovotics ISSN 2090-276X, e-ISSN 2090-7214
7. http://en.wikipedia.org/wiki/Roxxxy. Visited on 29 Dec 2012
8. Samani H, Cheok A, Ngiap F, Nagpal A, Mingde Q (2010) Towards a formulation of love in human-robot interaction. 19th IEEE International Symposium in Robot and Human Interactive Communication – Ro-Man 2010
9. Samani H, Cheok A (2010) Probability of love between robots and humans. In: IEEE/RSJ international conference on intelligent robots and systems, Taipei
10. Kuwamura K, Sakai K, Minato T, Nishio S, Ishiguro H (2013, August) Hugvie: a medium that fosters love. In: RO-MAN, 2013 IEEE. IEEE, pp 70–75
11. Mueller FF, Vetere F, Gibbs MR, Kjeldskov J, Pedell S, Howard S (2005, April) Hug over a distance. In: CHI'05 extended abstracts on human factors in computing systems. ACM, pp 1673–1676
12. http://cutecircuit.com/collections/the-hug-shirt/
13. Teh JKS, Cheok AD, Peiris RL, Choi Y, Thuong V, Lai S (2008) Huggy Pajama: a mobile parent and child hugging communication system. In: Proceedings of the 7th international conference on interaction design and children. ACM, pp 250–257
14. Brooks-Gunn J, Paikoff RL (1993) Sex is a gamble, kissing is a game": adolescent sexuality and health promotion. Promoting the health of adolescents: new directions for the twenty-first century. pp 180–208
15. Samani H, Parsani R, Rodriguez L, Saadatian E, Dissanayake K, Cheok A (2012) Kissenger: design of a kiss transmission device. In: Designing interactive systems conference – DIS 2012
16. Levy D (2006) A history of machines with sexual functions: past, present and robot. EURON Roboethics Atelier, Genova
17. Levy D (2006) Emotional relationships with robotic companions. EURON Roboethics Atelier, Genova
18. Levy D (2006) Marriage and sex with robots. EURON Roboethics Atelier, Genova
19. Levy D (2007) "Robot Prostitutes as Alternatives to Human Sex Workers" IEEE International Conference on Robotics and Automation, Rome. Reproduced as "The Ethics of Robot Prostitutes". In: Lin P, Abney K, Bekey G (eds) (2012) "Robot Ethics". MIT Press, Cambridge, MA, pp 223–231
20. Levy D (2012) The ethical treatment of artificially conscious robots. Int J Soc Robot 1:209–216
21. Sullins J (2012) Robots, love, and sex: the ethics of building a love machine. IEEE Trans Affect Comput 3(4):398–409

22. Amuda Y, Tijani I (2012) Ethical and legal implications of sex robot: an Islamic perspective. OIDA Int J Sustain Dev 3(6):19–28

23. Ziaja S (2011) Homewrecker 2.0: an exploration of liability for heart balm torts involving AI humanoid consorts. In: Mutlu B et al (eds) Proceedings of the international conference on social robotics (ICSR 2011), lecture notes in artificial intelligence (LNAI vol. 7072). Springer Verlag, Berlin, pp 114–124

24. Levy D (2012) When robots do wrong. Invited paper, Conference on Computing and Entertainment, Kathmandu, November 3rd–5th. Available at http://share.pdfonline.com/87cad18d73324e8fb2eaae1cddb60f77/Kathmandu_final_text_October31st.htm To appear in Lovotics, vol. 1

25. Russell A (2009) Blurring the love lines: the legal implications of intimacy with machines. Comput Law Secur Rev 25:455–463

26. Gates S (2013) Brazilian sex doll's virginity: bids for valentina's flower surpass $105,000. The Huffington Post, 7th Mar 2013

27. Hofilena J (2013) Japanese robotics scientist Hiroshi Ishiguro unveils body-double robot. Japan Daily Press, June 17 2013. Available at http://japandailypress.com/japanese-robotics-scientist-hiroshi-ishiguro-unveils-body-double-robot-1730686/

28. Torres I (2013) Japanese inventors create realistic female 'love bot'. Japan Daily Press, March 28, 2013. Available at http://japandailypress.com/japanese-inventors-create-realistic-female-love-bot-2825990/

29. Hugvie. www.geminoid.jp/projects/CREST/hugvie.html

30. Sumioka H, Nakae A, Kanai R, Ishiguro H (2013) Huggable communication medium decreases cortisol levels. Scientific reports, 3

31. Nishio S, Taura K, Ishiguro H (2012) Regulating emotion by facial feedback from teleoperated android robot. In: Ge et al (eds) International Conference on Social Robotics, LNAI 7621. pp 388–397

32. http://robotic.media.mit.edu/projects/robots/mds/headface/headface.html

33. DeSouza C (2013) Sexbots, ethics and transhumans. http://lifeboat.com/blog/2013/06/sexbots-ethics-and-transhumans

34. Fundawear Reviews (2013) www.fundawearreviews.com

35. Belford A (2013) That's not a droid, that's my girlfriend. The Global mail, February 21 2013. Available at www.theglobalmail.org/feature/thats-not-a-droid-thats-mygirlfriend/560/

Index