ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 4, July 2013

Utilizing Bimodal Emotion Recognition for Adaptive Artificial Intelligence

Alan Murphy, Dr. Sam Redfern
National University of Ireland Galway

***Abstract* - This paper presents an application of adaptive artificial intelligence and machine learning middleware to the development process of mainstream computer games. Every game has an intended purpose and intends to elicit a particular emotional response from its players. Horror games for example intend to elicit fear or fright, while other games may intend to elicit delight. Monitoring the emotions of a player during game play allows developers to assess whether their game is fulfilling its intended purpose. We initially investigate where emotional responses come from and then propose how they can be monitored. We also propose how utilizing our middleware in the beta testing stages of a game's development, can aid developers in extracting the information they require from players, in order to adapt game play to fulfill its intended purpose.**

***Index Terms* - Adaptive Artificial Intelligence, Emotion Recognition.**

I. INTRODUCTION

Artificial Intelligence (AI) can be described as the intelligence of machines or the design of intelligent agents that have the ability to perceive their environment and to take actions that maximize the chances of success. In terms of game play these agents generally take the form of an enemy bot or Non-Player Characters (NPCs) attempting to impede the player's attempts to accomplish his/her goal or task. Game developers are constantly intensifying their efforts to develop AI techniques that have the potential to have a powerful impact in several domains, including training, education and entertainment. Ram et al. (2007) describe AI as the effort of going beyond scripted interactions in order to enter a world of truly interactive systems that are "responsive", "adaptive", and "intelligent" [1].

Gaming although predominantly synonymous with entertainment can also refer to non-entertainment or 'serious' applications. These applications are designed with the primary purpose of applying simulation technologies to offer an immersive approach to the training and skills development of a workgroup. Such applications can include medical or military training applications that aim to train individuals or groups in a particular field. Whether the game's intended application is of a serious nature or entertainment nature, artificial intelligence will play a great role in going beyond simple scripted interactions. In attempting to assess player experience during game play - it is crucial that we take into account both the design of the world in which we are placing our players and also to assess the interactions they experience with other players or characters in the game.

These systems as described by Ram [1] constantly seek to learn about the player(s) and their personal or collective experiences during game play and use this information to transcend the pre-scripted intelligence provided by the developers. The result of this is a level of intelligence that will dynamically develop and provide a richer experience to the game player. All game development communities constantly seek new approaches, techniques and tools that will permit them to easily incorporate AI in their games. Examples of AI in mainstream games include Half Life [2] and F.E.A.R (First Encounter Assault Recon) [3]. When originally released in 1998 Half Life received critical acclaim for its influential and underlying artificial intelligence.

Its main innovations included

- An AI controlled security guard guided the player through some of the game levels
- First time use of squad AI in the later stages of the game
- Intelligence is incorporated smoothly into the storyline without the use of revolutionary technologies.

Figure 1 above shows the AI controlled security guard interacting with the player. Since Half Life (1998) games like F.E.A.R (2005) have intensified the level of game AI by incorporating enemies that are capable of cleverly interacting with a changing environment by finding cover behind tables, tipping bookshelves and opening doors. All of this scripted intelligence has a dramatic effect on a players experience during game play. Park et al. (2012) for example outline how playing either with or against scripted NPCs and the level of intelligence of the NPCs can

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 4, July 2013
have a profound impact on the enjoyment levels of the player [5]. Participants of the study in question portrayed how playing a game with a robot as an ally was more enjoyable and easier than playing against an intelligently scripted NPC. Research until now in this field of AI has focused on the enemies and focused only a small amount on the players' reactions to these enemies. Such a perspective could provide game developers with the player data they require in order to script the AI for such enemy NPCs in order to elicit the expected emotional response from players. By eliciting the expected emotional response, developers can assume that they are creating the expected experience for the player(s). This begs the question, why do people play games and what do they expect to experience?

Fig 1: Player interacting with security guard - Half Life [2]

Lazzaro (2004) asserts that people play games in general to change their internal experiences [5]. He describes how within any genre (First Person Shooter, Survival Horror…), every game generally has a storyline which could have many themes – it is this theme that will determine the experience the end user is expected to have. Lazzaro explains how most adults enjoy filling their heads with thoughts and emotions unrelated to those experienced in their workplace and that others enjoy the chance to test their abilities [5]. In order to establish whether people play games to feel emotions as well as to challenge, whether people modify games to feel differently and whether it is possible to build emotions into games, Lazzaros' studies concluded that there are four types of fun in games: 'hard', 'easy', 'serious' and 'people' fun [5]. The means in which the data was gathered however was largely based on questionnaires and interviews post game play, while little data was gathered from users during actual game play. This study gives our research a motive and a driving force to develop middleware which can be used as a development tool for developers in the beta testing stages of development. This middleware provides the capability of capturing a user's emotional response to a game or a scenario within a game, and therefore appropriately adjust the game's AI in real time, in order to veer closer towards the emotional response the game was designed to elicit from the player.

II. VIRTUAL GAMING ENVIRONMENTS

A. Designing Game Environments

The design of a game environment is pivotal to the experience of a player therein. A virtual gaming environment is the system which is designed to facilitate game play. When considering the design of these systems it is imperative to consider the end user context: Will this environment be intended to entertain the players? Will this be a virtual training or learning environment? The environment design will invariably differ depending on the nature of the application. In designing environments intended to elicit a certain emotion, it is important to take this into account in the environment design, in order to give it the appropriate meaning. In order for any game to be successful, intuitive negotiation with fellow players of mission related activities or task related activities in training applications is essential. In order to support this natural negotiation, tools must be put in place to support natural communication within the gaming application. Design of gaming environments must also incorporate the support for specific channels of communication. Foulger (2004) describes communication as "the process by which people…construct representations of meaning such that other people can interpret those representations" [6].

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 4, July 2013

When referring to game environment design, it is assumed that collaborative training games should support communication structures that will provide the bridges over the difficulties of communicating face to face vs. communication over a network [6]. The same can be said for games intended to entertain. In each case the messages transmitted between players, verbal or nonverbal – all contain emotional information which can potentially be used to inform adaptive AI scripts for game objects.

B. Emotion in Communication – Verbal and Non Verbal

In terms of our middleware tool it is important to establish what our raw data will be for our classification of emotion. Before we do this it is important to distinguish between verbal and nonverbal communication. Verbal communication or "oral communication" is the most obvious type of communication which takes place among human beings. As with any form of communication, verbal communication comes with its advantages and disadvantages. When communicating verbally, the communication speed can be controlled so as to allow for the delivery of points of information one by one, to make sure that each point is clearly communicated and understood before moving on to the next. This control can greatly increase the speed and efficiency of communication. The speaker can also be constantly aware as to whether or not the instruction or message to be communicated was clearly understood, or if the receiver is in a state of confusion. It is also true that verbal communication is much more precise than non-verbal cues. However it is important not to undermine the importance or power of nonverbal cues.

In face to face communications, there is much more information conveyed to a receiver other than just words. Nonverbal communication describes all those other means of conveying information. Verbal communication has been discussed in the previous section and it was described how it can be greatly complemented by nonverbal communication. Ross Buck and C. Arthur Van Lear (2002) categorize verbal communication as the "intentional" use of language, be that signed, written or spoken - and nonverbal communication as the "non-intentional" communication or underlying emotional states and meaning, through the processing of certain signals given "spontaneously" by the communicator [7]. Considering this, given that verbal communication is much more precise than nonverbal communication; our studies to date are mainly focused on the spoken aspect of verbal communication- analyzing intricacies like intonation and rhythm in speech. Once we have classified the emotion the player is conveying it then becomes our task to adapt the AI of the game appropriately.

C. *Emotionally Adaptive Artificial Intelligence*

When discussing adapting a game or a game object's artificial intelligence based on a players' emotion, we can deduce that the reason this has to be done is because either the game play has become too difficult or that the game is not eliciting the appropriate/intended emotion from the player. Emotionally adaptive gaming is a relatively new concept in the mainstream game AI although it has been a topic for researchers in the past few years. Higson (2009) developed one of the first ever emotionally adaptive games which consisted of a 'minesweeper'-like task and incorporated user feedback through taking pulse measurement from the fingertips [8]. Higson argues that in game play emotions can be looked at as a two dimensional state: valence and arousal where 'arousal is excitement or stress and valence is unhappiness to happiness scale' [8]. Based on this user input the game assesses how you are feeling and can adapt the AI accordingly e.g. if you are stressed and losing, then the game's AI adversaries could become less skilled.

The above approach essentially simplifies emotions down to one of two. Emotion however is in reality a much more subtle phenomenon. Robert Plutchik, Wallace Friesen, Carroll Izard and Eitaro Masuyama have made significant contributions to the study and classification of human emotions over many years of research [9]. In addition Paul Ekman is renowned as one of the most influential researchers in the field of emotional psychology. Such researchers all hold the common belief that emotions can be classified as being either a basic emotion or a primary emotion. Plutchik (1980) outlines eight primary emotions as anger, fear, sadness, disgust, surprise, anticipation, trust, and joy, whereas Ekman (1980) classifies the basic human emotions as: anger, happiness, sadness, fear, disgust and surprise [9][10]. These can be further broken into positive and negative emotions but for the purpose of a 'clever' AI system we argue that it is crucial to know the exact emotion of a user in order to appropriately adapt a game's AI. The change in AI can either take place during game play, or as we are suggesting can be used as a beta testing tool so that developers know they are eliciting the desired emotion for the game genre. Now we must question how we will classify these emotions?

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 4, July 2013

D. Classification of Emotion

There has been much research for many years into methods of extracting emotion from facial expressions, voice signals and sentiment analysis of text utilizing natural language processing techniques. The study of automated facial expression and emotion recognition spans a number of fields; the most important of these include neural networks for pattern recognition, image-processing and motion capture [11]. Much of the work conducted into the synthesis and classification of emotions through non-verbal signals has been conducted for application to the field of Human Computer Interaction (HCI) with the goal of creating natural human-like machines [12]. Bimodal Recognition Systems describe the use of classifying emotion by more than one strand [13]. As shall be described in our methodology we have designed a bimodal recognition system which will classify emotion based on a voice signal and also text obtained via a voice-to-text converter. What this provides is two systems attempting to solve the same problem, therefore, optimizing potential for increased accuracy.

III. METHODOLOGY

For our prototype recognition system a standard feed forward neural network with back-propagation training is being used. This network processes the audio data from players during game play.

A. Designing Game Environments

One of the primary concerns regarding the design and setup of our emotion recognition neural network concerns the representation of the inputs to the network, and the form our training data takes. The basic structure takes the form of a standard feed forward neural network and consists of three separate layers: an input layer; a hidden layer where our classification is performed, and a final output layer with six outputs to represent each of Ekmans' six fundamental emotions. The classified emotion is then determined by the most dominant of the six outputs. Figure 2 below shows an outline of our neural network setup.

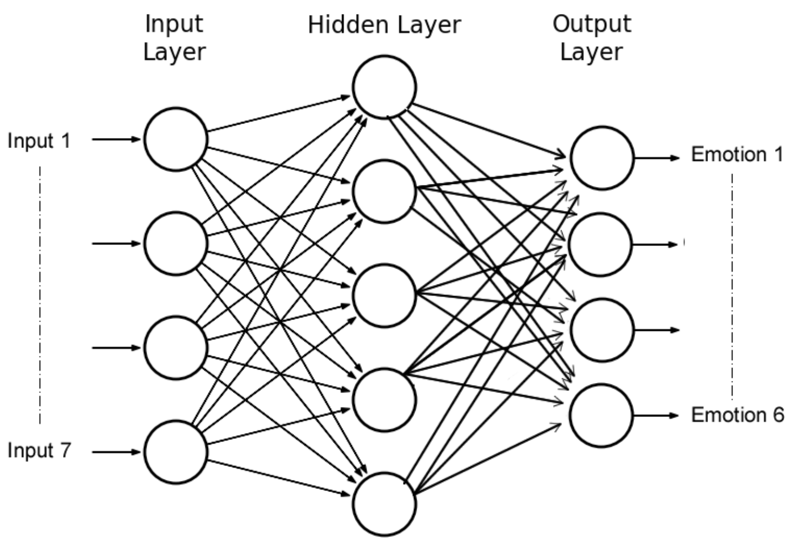

Fig 2: Diagram of our Neural Network Setup

B. Selecting Audio Parameters to Feed To the Network

In terms of selecting audio parameters to use as inputs to our network for initial training, there is much research literature which endeavors to derive the optimum set of parameters for emotional recognition through acoustic analysis. Schuller et al. (2009) perhaps provides the most informed and complete parameter list which should encompass all acoustic aspects of the emotions conveyed [14]. This list suggests the following seven acoustic parameters, and we use these as our base seven parameters, extracted using the open source acoustic toolkit 'OpenSmile' [15]:

- Raw Signal: Zero Crossing Rate

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 4, July 2013

- Signal energy: Logarithmic
- Pitch: Fundamental Frequency F0 in Hz via Cestrum and Autocorrelation
  Exponentially smoothed F0 envelope
- Voice Quality: Probability of Voicing
- Spectral Energy: Energy in bands 0-250 Hz, 0-650 Hz, 250-650 Hz, 1-4 kHz
- Mel-spectrum: Band 1-26
- Cepstral Coefficients: MFCC 0-12

The Zero Crossing Rate of a signal is the rate at which a signal oscillates from positive to negative, and when viewed on an oscilloscope portrays the raw signal. The absolute value of the area under this raw signal provides us with the Energy. The Pitch is based on the rate of vibrations creating the sound and can be used as a measure of tone. The Voice Quality parameter measures the quality of spoken utterances. This allows us to focus our analysis on areas of high voice quality for improved readings and improved emotion classifications. Spectral Energy describes how the energy of a signal is distributed with frequency over time. The final two parameters can be collectively described by the Mel Frequency Cepstrum (MFC). The MFC is the power spectrum of a signal and is based on a linear cosine transform of a log power spectrum on a nonlinear scale of frequency. The Cepstral Coefficients are coefficients that build up the MFC. The waveforms of a set of these parameters are shown below in Figure 3.

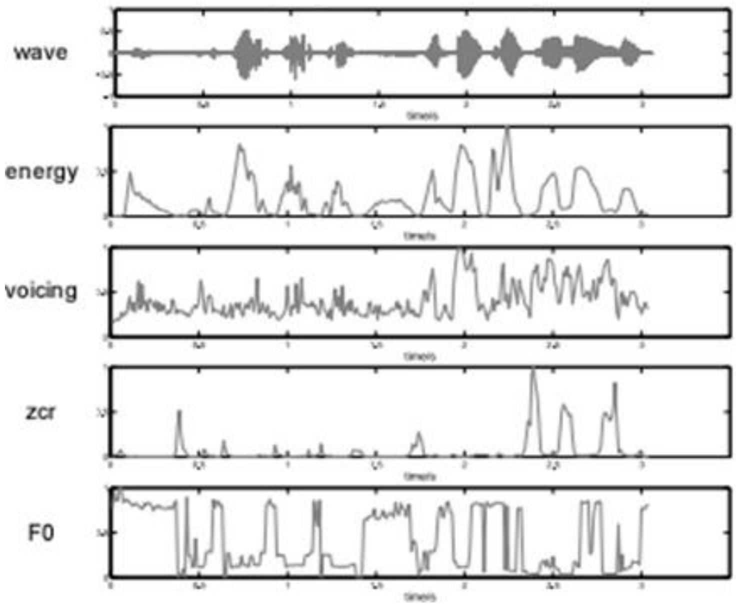

Fig 3: Samples of extracted speech parameters [16]

Through experimentation we found that the sampling rate of the input audio stream has a direct influence on the pitch extracted from the signal. By using a higher sampling rate of 48 kHz this ensures that the pitch reading is indeed accurate, and for lower sampling rates of 16 kHz pitch readings were obscure. The reason for this is that in order to calculate the pitch parameter we must rely on higher harmonics of F0. In the 48 kHz signal there are more of these harmonics than in a 16 kHz signal, so this implies that pitch extraction will be more robust using the higher sampling rate. Currently our work is focused around building the appropriate input schema to train our neural network. This in turn informs our work in further deriving the seven parameters as outlined by Schuller et al [14].

C. *Sentiment Analysis of Text*

As stated earlier we are in the process of developing a bimodal recognition system which incorporates a classifier for the text element of the speech. Given the preliminary nature of this aspect of the system are aiming to incorporate the text classifier into future work. Sentiment analysis through natural language processing has become a growing area of interest in the areas of opinion exploration, market analysis and also embodied agents. Our classifier is based on the ‘Wordnet-Affect’ database which is an extension of the ‘Wordnet’ natural language processing database [17]. ‘Wordnet’ presents a linguistic resource for a lexical representation of affective knowledge by assigning 'A-Tags' or affective tags to extracted words [17]. Through examining these A-tags -

moods, emotional states and general sentiment can be extrapolated. In order to prove our concept, deployment in a fully functional game or a game in its beta testing stages is required. Through monitoring conversations between players and logging their emotions in real time with our bimodal recognition system, developers can have a clear view of the emotional response being elicited from players and in turn know to alter the AI of the game accordingly if necessary. Our prototype could potentially also run as a real time adaptive AI middleware, adapting the AI in real time based on players conversations.

IV. CONCLUSION

To conclude we have presented an application of an adaptive artificial intelligence middleware to the development process of mainstream computer games. Whether a game's intention is to train or entertain, every game intends to elicit some kind of emotional response from its players. By monitoring a player's emotional responses in real time we aim to portray how our middleware can be of aid to developers in the beta testing stages of a game. This will provide useful player generated feedback in order to improve a game's intelligence scripts, adapt a games' intelligence in real time in order to move toward eliciting the desired emotional response.

REFERENCES

[1] Ram, Ashwin., Ontanon, Santiago & Mehta, Manish., (2007). Artificial Intelligence for Adaptive Computer Games. Twentieth International FLAIRS Conference on Artificial Intelligence, American Association of Artificial Intelligence.

[2] Half Life, Valve, 1998.

[3] F.E.A.R. First Encounter Assault Recon, Monolith Productions, 2005.

[4] Park, Eunil., Ki Joon, Kim., Shyam Sundar, S.,del Pobil, Angel P., (2012). Online Gaming with Robots vs. Computers as Allies vs. Opponents, ACM/IEEE international conference on Human-Robot Interaction, 205-206.

[5] Lazzaro, Nicole., (2004), Why We Play Games: Four Keys to More Emotion Without Story, Player Experience Research and Design for Mass Market Interactive Entertainment, 510-658.

[6] Foulger, Davis. (2004), "An Ecological Model of the Communication Process", Brooklyn College/CUNY, Retrieved: 4 February, 2013, http://davis.foulger.info/papers/ecologicalModelOfCommunication.htm

[7] Buck, R. and VanLear, C. A. (2002), Verbal and Nonverbal Communication: Distinguishing Symbolic, Spontaneous, and Pseudo-Spontaneous Nonverbal Behavior. Journal of Communication, 52: 522–541.

[8] Higson, Irene.,(2009), "Emotionally Adaptive Gaming", Accessed: 4 February 2012, http://www.physics.org/featuredetail.asp?id=42.

[9] Plutchik, R., (1980), "Emotion: A Psychoevolutionary Synthesis", New York: Harper & Row, ISBN 0-06-045235-8.

[10] Ekman, Paul, (1980), "The Face of Man: Expressions of Universal Emotions in a New Guinea Village", Garland STPM Press.

[11] Chibelushi, C. C., Bourel F., (2002), "Facial Expression Recognition: A Brief Tutorial Overview".

[12] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G.., Kollias, S., Fellenz, W., Taylor, J.G..,(2001),"Emotion recognition in human-computer interaction", Signal Processing Magazine, IEEE, 18(1), pp. 32-80.

[13] Chuang, Ze-Jing., Wu Chung-Hsien.,(2004) Multi-Modal Emotion Recognition from Speech and Text, Computational Linguistics and Chinese Language Processing, Vol:9 No:2, 45-62.

[14 Schuller, B.; Vlasenko, B.; Eyben, F.; Rigoll, G.; Wendemuth, A.; , (2009), "Acoustic emotion recognition: A benchmark comparison of performances," Automatic Speech Recognition & Understanding, pp.552-557.

[15] Florian Eyben, Martin Wöllmer, Björn Schuller, (2010), "openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor", Proc. ACM Multimedia (MM), ACM, Firenze, Italy.

[16] Eyben, F.; Wollmer, M.; Schuller, B., "OpenEAR — Introducing the munich open-source emotion and affect recognition toolkit," Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference pp.4.

[17] Strapparava, Carlo., Valitutti, Alessandro., (2004), "WordNet-Affect: an Affective Extension of WordNet", Proc. Intl. Conf. Language Resources and Evaluation, vol. 4, pp.1083 -1086 2004.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 4, July 2013

AUTHOR BIOGRAPHIES

Alan Murphy attended the Institute of Technology in Athlone, Ireland where he studied for a B.Sc. in Computer and Software Engineering (2008). He followed this by studying in the National University of Ireland, Galway where he obtained a M.A in Digital Media (2009). He then worked for one year at web application and database development for a Galway based Production Company (2010). Alan has since returned to full time education in NUIG where he is currently studying for his PhD in Information Technology.

Sam Redfern attended the National University of Ireland, Galway where he studied for a B.A in English and Archaeology (1992) followed by a M.Sc. (1994) and PhD (1998) in Information Technology. He has worked as a lecturer in Galway since 1996, and has since published in areas of digital image processing, various types of A.I, graphics, collaborative virtual environments and serious games. He has been an independent game developer in his spare time since 1984, with games published on the BBC Micro, Amiga, PC, Mac, iPhone and Android.