Inducing and Measuring Emotion through a Multiplayer First-Person Shooter Computer Game

Paul P.A.B. Merkx[1], Khiet P. Truong[1] and Mark A. Neerincx[1,2]

[1] TNO Human Factors, P.O. Box 23, 3769ZG Soesterberg, The Netherlands
[2] Delft University of Technology, P.O. Box 5031, 2628CD Delft, The Netherlands
{paul.merkx,khiet.truong,mark.neerincx}@tno.nl

Abstract. To develop an annotated database of spontaneous, multimodal, emotional expressions, recordings were made of facial and vocal expressions of emotions while participants were playing a multiplayer first-person shooter (fps) computer game. During a replay of the session, participants scored their own emotions by assigning values to them on an arousal and a valence scale, and by selecting emotion category labels (e.g. 'happiness' and 'frustration'). The fps-game proved to be a successful evoker of a diversity of emotions. The annotation results revealed interesting insights in current models of emotion. While the two-dimensional arousal-valence space is usually described as circular, we found a V-shape pattern of reported arousal and valence values instead. Furthermore, correlations were found between emotion characteristics and user-specific preferences about games. The recorded data will be used in future research concerning automatic and human recognition of emotions.

1 Introduction

Emotions and automatic emotion recognition increasingly play important roles in designing and developing computer games [1, 3]. Automatic emotion recognition is an emerging technology that can provide an enormous boost to the entertainment industry. Imagine an adaptive computer game that adjusts its difficulty level to the observed frustration level of the user [2]. Or imagine a speech-enabled computer game that responds to affective voice commands [6]. In order to develop affective videogames, the system first needs to be able to sense the player's emotional state in a least obtrusive manner. Physiological measures such as heart rate or skin response can be relatively good predictors of a human's emotional state but these are usually measured through special equipment that is attached to the body, which can be experienced as obtrusive by the player. Other channels through which emotions can be expressed are voice and facial expressions. Vocal and facial expressions of emotions can be relatively easily registered by a camera and a microphone which are usually perceived as less obtrusive. Hence, we focus on measurements of vocal and facial expressions in this study.

There is a vast amount of literature available on the automatic recognition of vocal and facial expressions of emotions (e.g., [4, 5]). However, it is difficult to compare these studies to each other and draw conclusions about the performances of the recognizers with respect to its applicability in a real-world situation. A reason for this, among others, is that there is no agreement on how to represent or describe emotions. Further, most of these studies are based on typical, full-blown emotions that were acted out by actors. We rarely encounter these typical extremities of emotions in real-life situations: in real-life, we tend to adhere to the unwritten social conversational rules and express more subtle emotions. One of the obstacles in emotion research is the lack of annotated (i.e., information about *what* type of emotion occured *when*) *natural* emotion data. For the development of automatic emotion recognition applications that employ machine learning techniques, a large amount of natural emotion data is needed to train and test the emotion models. Several eliciation methods have been employed in the past to evoke natural, spontaneous affective responses. For example, efforts have been made to elicit spontaneous emotions by showing movies or pictures (e.g., [7, 8]), by interacting with spoken-dialogue systems or virtual characters [9–11] and by playing games [12–15]. Since most of the games are designed to evoke emotions and we think that games can trigger a relatively broad range of different types of emotions, we used a multiplayer first-person shooter computer game to elicit affective responses and to study what type of emotional experiences and expressions are associated with this type of computer game.

In this paper, we present our first results of this data collection effort and illustrate how computer games can be used to elicit affective responses. Vocal and facial expressions of the players were registered during gameplay and, subsequently, were evaluated on emotion by the players themselves. Two different annotation methods were used. In the first method, participants were asked to rate their own emotions on continuous scales of arousal (active vs. passive) and valence (positive vs. negative). In the second method, a categorical description of emotions was adopted and participants were asked to choose between a number of predefined emotion labels. They were also given the possibility to give their own emotion label if there was no appropriate label present.

This paper is organized as follows: Section 2 will present a short summary of the most important topics in current emotion research. Section 3 will describe how the experiment was accomplished, while Section 4 will outline its results. Finally, Section 5 will discuss how the results can be interpreted and used for future research.

2 Emotion theory

There is an ongoing debate about the question how emotion can be defined and represented, especially concerning categorical and dimensional (or continuous) approaches to emotion. Research that supports the categorical approach to emotion usually asserts the existence of a number of basic emotions that are universally expressed and recognized (e.g., [16, 17]). The best known list of basic emotions is often termed the 'Big Six': happiness, sadness, fear, anger, surprise and disgust [22]. Corresponding with the categorical approach, emotions can be annotated by assigning category labels to them. The following twelve emotion labels were used in our experiment: happiness, surprise, anger, fear, disgust, excitement, amusement, relief, wonderment, frustration, boredom and malicious delight. This list of labels was based on basic emotions and typical game-related emotions (derived from [1]).

Research that supports the dimensional approach models emotion as a two- or three-dimensional space (e.g., [18, 19]). As an annotation method, participants can be asked to define a point in space that corresponds to a particular emotion. This point can be defined as the intersection of two or more values. Since consistently identifying a third dimension (e.g. tension, control or potency) proves to be difficult [24], two dimensions are usually used, which are labeled as arousal (ranging from 'calm' i.e., low arousal to 'excited' i.e., high arousal) and valence (the degree to which the pleasantness of an emotional experience is valued as negative or positive, thus ranging from highly negative to highly positive). An example of such a method is the Feeltrace annotation tool (see [19]). Although some agreement exists that reported values in arousal-valence space form a circular pattern, results in [25] describe annotation tasks where participants had to watch pictures or listen to sounds and where reported arousal-valence values formed a V-shaped pattern in the two-dimensional space.

In our experiment, the name of the arousal scale was changed to 'intensity'. This was done to make the task more accessible for naives, because using the old term might have yielded less consistent results due to the misunderstanding of the exact meaning of the term. In addition, there is no real appropriate Dutch translation of 'arousal' with respect to emotion annotation.

The next section describes how we applied the arousal-valence model and the category labels for the acquisition of an annotated emotion database.

3 Method

An experiment was performed in order to elicit emotional responses from participants who were invited to play a computer game and to carry out an experimental task.

3.1 Participants

Seventeen males and eleven females participated in the experiment with an average age of 22.1 years and a standard deviation of 2.8. They were recruited via e-mail under a selection criterium of age (between 18 and 30 years). People suffering from concentration problems were not allowed to participate in the experiment. In addition, participants were asked to bring a friend or relative in order to have a teammate who they were already acquainted with prior to the experiment. The reason for this is the fact that people report and show more arousal when playing with a friend compared to playing with a stranger [20], especially when friends are playing in the same room [21]. A compensation was paid to all participants, sometimes in combination with a bonus (which was rewarded to the players of the two teams with the highest scores in the game and to the two teams with the best collaborating teammates). The function of these bonusses was to keep the participants motivated and to encourage them to talk while playing since one of our goals was to elicit and record vocal expressions.

3.2 The game

Participants played in two teams of two players. The game they played was the first-person shooter Unreal Tournament 2004 by Epic Games, which has the possibility for players to play the game with and against each other through a Local Area Network (LAN). Unreal Tournament 2004 offers diverse multi-player game modes, from simple 'deathmatches' to rather complex mission-based modes. Since at least some participants were expected to be unfamiliar with the game, the game mode 'Capture the flag' was selected to be used in the experiment. This game mode has a low learning curve, but nevertheless invites strategy talk. A very small 3D world (called '1on1-Joust' in the game) was selected for the participants to play in, in order to evoke hectic situations in which (tactical) communication and frustrations would easily arise. For both teams the goal of the game was to capture the other teams flag as many times as possible while defending their own flag.

At any time in the game (for example, in less exciting periods in the game) the experimenter was able to generate unexpected in-game events to keep players interested and surprised. Some examples were the sudden appearence of monsters, the sudden ability to fly, an increasing or decreasing game speed and unexpected problems with the gameplay. These events were evoked at an average frequency of one event per minute using Unreal Tournament's build-in cheatcode list.

3.3 Apparatus

Four PC's (connected via a LAN) were used as game computers, while webcams and microphones were connected to four laptops, one for each participant. Video recordings were made using Logitech Quickcam Sphere webcams. In-game events were captured by creating a screenshot of the game every second. These screenshots were concatenated to one single avi file.

3.4 Procedure

In short, the time schedule for the experiment was as follows:

- General instruction (15 minutes)
- Training session: getting used to the game (10 minutes)
- Instruction about the computer task followed by a short break (20 minutes)

- Training session: getting used to the computer task (20 minutes)
- Session 1a: playing the game (20 minutes)
- First questionnaire followed by a break (25 minutes)
- Session 1b: computer task with a ten-minute break halfway. (50 minutes)
- Long break (40 minutes)
- Session 2a: playing the game (20 minutes)
- Second questionnaire followed by a break (25 minutes)
- Session 2b: computer task with a ten-minute break halfway. (50 minutes)

After reading a general instruction about the experiment, participants received a training session in which they could get used to the game rules and gameplay. During this session, which took ten minutes, the experimenter provided the players with some additional guidelines about the game, helping them to play and enjoy the game better. Beforehand, the experimenter reminded the players about the bonusses they could receive and stimulated them to discuss their strategies and to get a score as high as possible. Throughout the training session, recordings were made of players' faces and speech and of the in-game events.

Subsequently, participants received instructions about the experimental task they had to perform after each game (see Section 3.6). A thorough twenty-minute training session allowed them to ask questions about the task and to be corrected by the experimenter if necessary.

After they finished their training, this same procedure of playing and annotating was repeated two times while play time and consequently annotation time were twice as long. The schedule allowed for long breaks of twenty to thirty minutes between sessions and shorter breaks within sessions. Participants were also asked to fill in two short questionnaires between sessions (see Section 3.5).

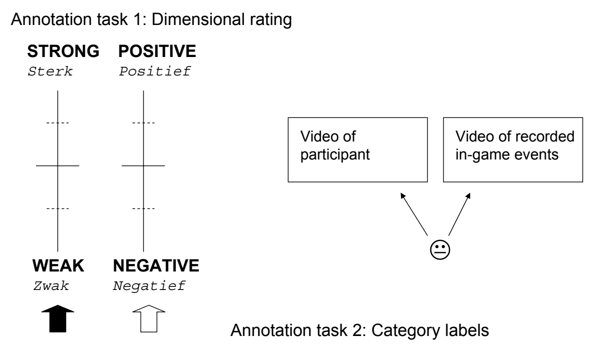
3.5 Questionnaires

Participants performed a bipolar Big-Five personality test (see [23]) and a gaming-related questionnaire. The gaming-related questionnaire asked among others:

1. How much do you like to play games in general?
2. How much do you like to play first-person shooters in general?
3. Did you like to play the game in the experiment?
4. How much time do you play computer games in general?
5. How much time do you play computer games together with other people, for example using the internet?

3.6 The annotation task

The recorded material was presented to the participants only twenty minutes after playing, including continuous recordings of the in-game events. While watching and hearing their own facial and vocal expressions and the captured video



Annotation task 2: Category labels

Happiness *Blijdschap*	**Excitement** *Opgewondenheid*	**Boredom** *Verveling*	**Anger** *Boosheid*
Amusement *Amusering*	**Relief** *Opluchting*	**Surprise** *Verbazing*	**Frustration** *Frustratie*
Malicious delight *Leedvermaak*	**Wonderment** *Verwondering*	**Fear** *Angst*	**Disgust** *Walging*

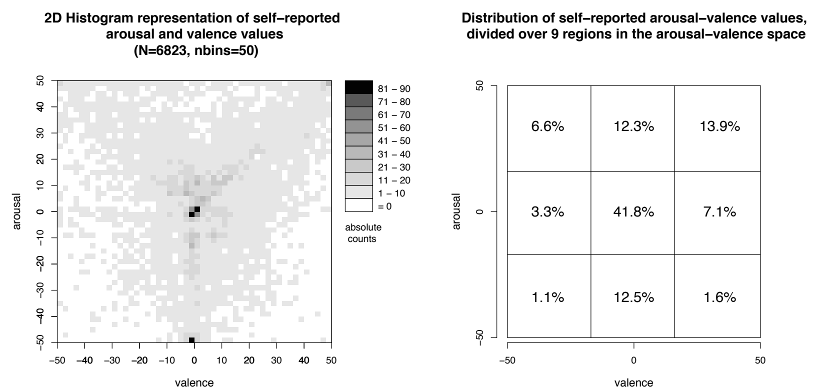
Fig. 1: *The participant annotates his/her own emotions by watching his/her own video and the captured video stream from the in-game events. The participant performs two different annotation tasks. The Dutch translations that we used are in italics.*

stream of the game, participants were asked to remember which emotions they experienced and how strong those emotions were at exactly that moment in the game. A two-part computer task was created for the annotation process. The first part asked participants about the intensity (ranging from weak to strong) and valence (reaching from negative to positive) of their emotions on two continuous scales, presented vertically and next to each other (as can be seen in Figure 1, together with Dutch translations, since the participants were Dutch). It is presumably easier for naive participants to perform the annotation on one scale at a time rather than evaluating both scales at the same time in a two-dimensional space, as is the case with the Feeltrace method [19]. Every ten seconds, an arrow indicated that a value on the intensity scale had to be selected, i.e. that a point on the first axis had to be clicked. Subsequently, a second arrow appeared, indicating that a value on the valence axis had to be clicked.

During the second part of the annotation task, participants watched and listened to the same material again and were asked how they would label their experienced emotions. This time, they were able to define the start and end points of the emotional moments themselves, activating or de-activating labels if a corresponding emotion occured or disappeared respectively. As previously mentioned, we selected twelve labels from which participants had to choose. Although the number of labels was kept as small as possible in order not to complicate the task for the participants, we expected that these labels would cover the whole range of emotions that could occur. In addition, participants had the possibility to verbally mention a term that they found more suitable for a certain emotion than the labels that the computer task contained. Figure 1 gives an overview of both the first part and the second part of the annotation task.

4 Results

Figure 2a gives a global overview of the reported values of all participants in the continuous annotation task. Darker areas contain more values than lighter areas. Figure 2b shows the percentage of reported values for nine subareas of the two-dimensional space. As can be derived from figures 2a and b, most values were selected from the center area of the two-dimensional space, while the upper-right corner with its high arousal and positive valence was another frequently visited region. In cases of low arousal, valence was nearly always analyzed as close to neutral, while high arousal was accompanied by more diverse valence values.

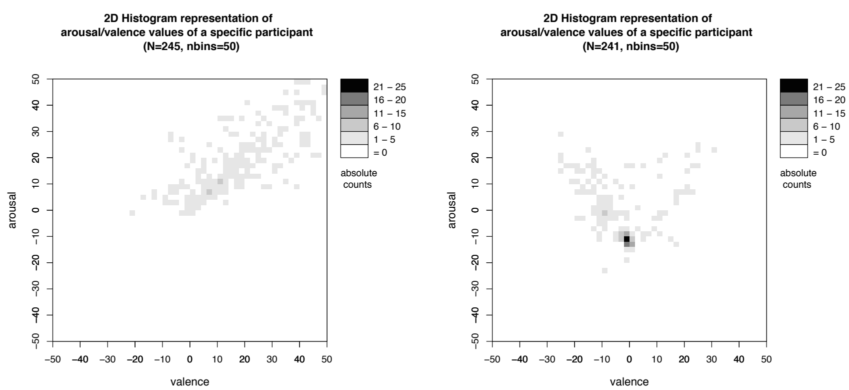


(a) 2D-histogram (b) Density divided over 9 regions

Fig. 2: *Graphs representing the density of reported arousal-valence values of all participants, plotted in an arousal-valence space.*

We can observe in Figure 2 that the reported values follow a certain pattern, which can be described as a 'V-shape'. This is not entirely surprising since it is difficult to imagine an extremely negative or positive vocal or facial expression with weak intensity. Therefore, we calculated correlations between intensity and valence in order to see whether this observation can be supported. A significant quadratic relationship between arousal and valence was found with r=.47 (significant at $p<.01$). Furthermore, in the positive valence half of arousal-valence space, a linear relationship between arousal and valence with r=.55 ($p<.01$) was found. A corresponding negative correlation between arousal and valence was found in the negative valence half, with r=-.46 ($p<.01$). These correlation figures suggest that higher intensity values are reported if valence is more extreme to either the positive or the negative side.

Fig. 3: *Movie stills of registered emotions from four participants.*



(a) an enthusiastic player (b) a not so enthusiastic player

Fig. 4: *2D-histograms representing the density of reported arousal/valence values of two examples of different types of players, corresponding to the upper left and right players from Figure 3 respectively.*

The diversity of emotions can also be attributed to the fact that different types of players participated in the experiment. Figure 3 shows some examples of movie stills of registered emotions from four participants. Figures 4a and b show plots of reported values in the two-dimensional space of an enthusiastic and an often frustrated not so enthusiastic player, corresponding to the upper left and right players from Figure 3 respectively (who also reported in the gaming-related questionnaire they liked and disliked the game respectively). The plots show again an overview of the reported values. It is clear that the reported values of these players still fall within the V-shape pattern, despite of individual differences between the players.

There seems to be a relation between reported valence values and the amount of everyday life multiplayer playing time (Pearson's r=.49; significant at p<.05) and on the question whether players like the first-person shooter genre in general (Spearman's r=.50; p<.05). The reported arousal values seem to be related with both the questions 'do you like to play games in general' (Spearman's r=.50; p<.05) and 'did you like to play the game in this experiment' (Spearman's r=.45; p<.05). In all of these cases, average arousal or valence values were higher if questions were answered more positively or if players played more often in everyday life.

5 Conclusions and discussion

The most important aim of this study was to collect annotated natural emotion data through the use of a first-person shooter computer game and two different types of annotation methods. The first-person shooter proved to be a good elicitation tool for emotion. The game evoked a lot of different and intense emotions (ranging from negative to positive and from low to high arousal), probably caused by its intense atmosphere and the fact that the game was played by a group of four people who were all motivated to receive a reward. The game also evoked a lot of vocal expressions (in addition to facial expressions), probably caused by the fact that players were rewarded for good collaboration and the necessity of discussing strategies in order to perform better in the game. Making recordings proved to be relatively easy while participants were playing: they sat in front of their monitor while the webcam and the microphone registered all the expressions. In general, using a game as elicitation tool offers the opportunity to register relatively natural emotions in a somewhat laboratory environment: by playing computer games, participants become more immersed in a virtual world and pay less attention to their environment.

A first look at the annotation results offers some interesting insights considering the arousal-valence space. Our acquired arousal-valence values show that, while reported values in arousal-valence space are usually described as forming a circular pattern (e.g. [18, 19]), participants clearly followed a V-shape pattern while they were annotating their emotions. The V-shape pattern implies that participants interpreted a very negative or very positive emotion automatically as being accompanied by a high level of arousal and a more neutral emotion as having a lower arousal level. Note that the pattern of the reported arousal and valence values can be dependent on the type of game. However, our finding is in line with [25], which was interestingly the result of a very different emotion elicitation experiment. We can imagine that a very positive or very negative emotion is rather difficult to associate with low arousal. Future research might therefore focus on the validity of a circular or rectangular arousal-valence space. It might be interesting to investigate whether different emotion evokers yield different patterns of values in arousal-valence space. Further data analysis will reveal where naive participants think the emotion category labels from the second annotation task should be placed in the arousal-valence space.

Self-reported valence values were increasingly higher for players who spend more time on playing games with other humans in their spare time. This fact is a nice addition to what was concluded in [20], namely the fact that self-reported valence is more positive when people play against another human compared with playing against the computer. Hence, it appears that 'The people factor' is an important evoker of positive emotions in games, as was already mentioned in [1]. Furthermore, self-reported valence was also more positive if players liked the first-person shooter genre better. Arousal values were increasingly higher if players liked playing games in general or playing the game during the experiment. So it seems that in general, players experience more positive and/or more aroused emotions if they like a game or the multiplayer aspect of a game better.

In a follow-up study, participants will evaluate and annotate a selection of the recorded material in different modalities (visual and auditive). These annotations will be compared with the annotations made by the players themselves, trying to answer questions like 'How good can humans recognize emotions in different modalities (visual and auditive), with or without contextual information (the captured video stream of the in-game events) from which the emotions emerged?' and 'Are arousal and valence levels equally difficult to recognize in different modalities or do differences exist between them?'. Furthermore, the recorded data will be used for multimodal automatic emotion recognition research. Subsequently, we can investigate how human emotion recognition compares to automatic emotion recognition.

6 Acknowledgements

This study is supported by the Dutch BSIK project MultimediaN.

References

1. Lazarro, N.: Why we play games: 4 keys to more emotion without story. Presented at the 2004 Game Developers Conference (2004)
2. Gilleade, K.M., Dix, A.: Using frustration in the design of adaptive videogames. Proceedings of ACE, Advances in Computer Entertainment Technology (2004) 228–232

3. Gilleade, K.M., Dix, A., Allanson, J.: Affective videogames and modes of affective gaming: Assist me, challenge me, Emote Me (ACE). Proceedings of DIGRA Conf. (2005)
4. Ververidis, D., Kotropoulos, C.: Emotional speech recognition: Resources, features, and methods. Speech Communication **48**(9) (2006), 1162–1181
5. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: The state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(12) (2000), 1424–1445
6. Jones, C.M., Sutherland, J.: Creating an emotionally reactive computer game responding to affective cues in speech. Proceedings of the 19th British HCI Group Annual Conference (2005)
7. Lang, P.J.: The emotion probe - studies of motivation and attention. American Psychologist **50** (1995), 371–385
8. Wagner, J., Kim, J., Andre, E.: From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification. Proceedings IEEE ICME International Conference on Multimedia and Expo (2005), 940–943
9. Ang, J., Dhillon, R., Krupski, A., Shriberg, E., Stolcke, A.: Prosody-based automatic detection of annoyance and frustration in human-computer dialog. Proceedings ICSLP (2002), 2037–2040
10. Batliner, A., Fischer, K., Huber, R., Spilker, J., Noth, E.: How to find trouble in communication. Speech Communication **40** (2003), 117–143
11. Cox, C.: Sensitive artificial listener induction techniques. Presented to HUMAINE Network of Excellence Summer School (2004), [Online Available: http://emotionresearch.net/ws/summerschool1/SALAS.ppt259]
12. Kim, J., Andre, E., Rehm, M., Vogt, T., Wagner, J.: Integrating information from speech and physiological signals to achieve emotional sensitivity. Proceedings Interspeech (2005), 809–812
13. Johnstone, T.: Emotional speech elicited using computer games. Proceedings ICSLP (1996), 1985–1988
14. Yildirim, S., Lee, C.M., Lee, S.: Detecting politeness and frustration state of a child in a conversational computer game. Proceedings Interspeech (2005), 2209–2212
15. Wang, N., Marsella, S.: Evg: an emotion evoking game. Proc. 6th International Conference on Intelligent Virtual Agents (2006)
16. Ortony, A., Turner, T.J.: Whats basic about basic emotions? Psychological Review **97**(3) (1990), 315–331
17. Ekman, P. (1999).: Basic emotions. In Dalgleish, T., Power, M.: Handbook of cognition and emotion (1999), 45-60
18. Russell, J.A.: A circumplex model of affect. Journal of personality and social psychology **39**(6) (1980), 1161–1178
19. Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schröder, M.: 'Feeltrace': An instrument for recording perceived emotion in real time. Proceedings of the ISCA Workshop on Speech and Emotion (2000), 19–24
20. Ravaja, N., Saari, T., Turpeinen, M., Laarni, J., Salminen, M., Kivikangas, M.: Spatial presence and emotions during video game playing: Does it matter with whom you play? Proceedings Presence (2005), 327–333
21. Halloran, J., Rogers, Y., Fitzpatrick, G.: From text to talk: Multiplayer games and voiceover IP. Proceedings of Level Up: First International Digital Games Research Conference (2003), 130–142
22. Cowie, R., Cornelius, R.R.: Describing the emotional states that are expressed in speech. Speech Communication **40** 2003, 5–32

23. Goldberg, L.R.: The development of markers for the Big-Five factor structure. Psychological assessment **4** 1992, 26–42
24. Scherer, K.R.: What are emotions? And how can they be measured? Social Science Information **44**(4) 2005, 695–729
25. Bradley, M.M., Lang, P.J.: Affective reactions to acoustic stimuli. Psychophysiology **37** 2000, 204–215