Proceedings, The Eighth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment

Enhancing the Believability of Character Behaviors Using Non-Verbal Cues

Neesha Desai and **Duane Szafron**
University of Alberta
Edmonton, AB, T6G 2E8, CANADA
neesha | dszafron @ ualberta.ca

Abstract

Characters are vital to large video game worlds as they bring a sense of life to the world. However, background characters are known to rarely exhibit any sign of motivated behavior or emotional state. We want to change this by assigning these characters emotions that can be identified through their non-verbal behavior. We feel the addition of emotion will allow players to feel more connected to the game world and make the game world more believable. This paper presents the results of an experiment to test two ways of conveying emotion: 1) through a character's gait and 2) through a character's interactions with the game world. Results from the experiment suggest that a combination of gait and interactions is the most effective method to convey emotion.

Introduction

Complex video game worlds, such as the ones in story-based games like Skyrim (Bethesda 2011) and Mass Effect 3 (BioWare 2012), require hundreds, if not thousands, of background characters. These characters may be friends or foes of the player's character. They may have a pivotal role in the story, a small role in a side quest, or no purpose outside of bringing a sense of life to the game world.

The characters that have a pivotal role in the story are usually scripted, given a name, dialogues, and follow a series of actions that allow them to exhibit behaviors and, in some games, express emotions. The rest of the characters, those that are rarely (if ever) used in the story, usually have few, if any, scripts attached and exhibit few signs of motivated behavior. These are the characters that are often found standing around staring at walls. If they have any dialogue, they usually repeat the same set of phrases. These characters do not do much and rarely exhibit any sign of motivated behavior or emotion through their actions. One example is the set of crew members on Commander Shepard's spaceship in Mass Effect 3. As Shepard explores the ship, the crew members are spread about looking like they are in conversation with others or controlling some aspect of the spaceship. However, when approached and/or watched, it becomes apparent that they are repeating the same sequence of actions/animations and that there is no dialogue actually being exchanged. Even walking into these characters does not generate a response.

As well as lacking believable behaviors, background characters rarely act (or react) in a way that indicates they have underlying goals or purposes behind their actions. We explore the notion that if a player projects an emotion onto these characters, the player will subconsciously or consciously ascribe motivated behaviors to these characters that will increase believability. Currently, characters usually only present emotion through their dialogue lines and/or actions and animated facial expressions in cut scenes. As video games develop more complicated and deeper story lines, and as the serious games genre grows, the need for characters to express believable and identifiable emotion will increase.

Much research has been done on how people identify emotion. Most of this research has focused on how our faces express emotion, although there is a smaller body of work on detecting emotion from other forms of non-verbal behavior (Wallbott and Scherer 1986). While facial animations are able to elicit emotional responses from viewers, they require a lot of computational power (Cao et al. 2005). Most facial animations involve small movements that are difficult to identify without a close-up. In video games, facial animations are used mainly in cutscenes, when the game can use close-ups to present faces at a higher resolution.

Video game characters have the restriction of limited resources. There is so much being computed during each frame of a game, that adding additional computations is expensive and often impossible. This means that any technique used to add emotion to a character must require very limited resources and should rely, primarily, on the existing game content. Uniquely scripting each character requires a lot of time and money. And as the game worlds grow, so do the number of characters that reside within.

We believe that the display of emotion by characters will improve the ability of games to create stories that the players can connect with, add depth to the story, and provide another level of believability. Emotional characters will be especially important for serious and training games. Specifically, we are trying to foster an illusion of emotion-driven character behavior by using an assigned emotion to modify how characters interact with the game world. We are relying on the character's non-verbal behavior to present their emotional state to the player. This includes manipulating their stride, walking path, and interactions with objects in the game world.

Artificial intelligence in games often revolves around creating agents that can produce or mimic human-style play and problem solving methods. In some games it centres around creating characters that players may believe to be controlled by humans. Our research attempts to create artificial agents that are perceived as being under the influence of particular human emotions. The results of our experiment show that participants can accurately and easily identify the emotion of these characters, which will allow these characters to be seen as more 'human-like' or 'believable'.

Related Work

Animation

Animators have long used non-verbal behavior to convey emotion. This can be seen in any Disney or Pixar animated movie. The Pixar short film *Luxo Jr.*, about a 'young' lamp's interaction with a rubber ball (Pixar 1986) is able to convey the lamp's joy and sadness, even though the lamp does not have a face or ability to speak. Ken Perlin (2011) has also done research in expressing emotion, one example being the simple polygon in *Polly's World*. However, one characteristic these examples have in common is that they rely on exaggerated movements. This works well when the object is not supposed to be human-like, just express some behaviors that are creature-like. However, a serious game showing exaggerated behavior is likely to push negative stereotypes and prevent players from connecting. We want to create non-verbal behavior that is similar to real life, possibly using slight exaggeration for effect, but, not overly exaggerated.

When prototyping a serious game, we found that basic character animations may be perceived differently by psychologists and their patients (Desai et al. 2011). The prototype was created using Neverwinter Nights, using a modern art kit. For example, when shown a conversation scene with characters displaying the stock idle/talk animations, the psychologists stated that the characters had a very aggressive stance. They doubted that their depressed patients would be able to identify with the characters in the serious game. Neverwinter Nights was most likely* not trying to capitalize on people's ability to recognize and identify non-verbal behavior. In this case, the wrong message about emotion inferred from the animations made Neverwinter Nights un-useable for prototyping, beyond possible blocking of a scene.

Identifying emotion through gait

Roether *et al.* (2009) conducted a study in which they developed five different walking animations to represent the four basic emotions (happy, sad, angry, and afraid (Ortony and Turner 1990)) and neutral. These animations were placed on a simple human form similar to an artist's drawing mannequin. The results of their user study showed that people could generally identify emotion based on gait. We have replicated and extended their study, using a video game character with more believable features.

Emotion through non-verbal behavior

Examining whether emotion can be recognized from nonverbal cues can be traced back decades (Wallbott and Scherer 1986), but it has most commonly been tested with human actors. More recently, Seif El-Nasr and Wei (2008) looked at creating non-verbal behavior models. Seif El-Nasr and Wei had animators create an office scene containing two characters discussing one character's job search. The animators were asked to make a series of scenes. In each scene, each character was provided a short description, such as 'to impress someone' or 'to be thought normal'. They found that for some of the descriptions, the animators expressed them differently depending on which actor they were applying it to. This suggests that, for the animators, the character's identity or perceived importance affected how they thought that character would express a given state-of-mind.

Our method

We examined two techniques that will allow characters to display emotions that players can easily and accurately identify. We wanted to take into account that video games cannot always show close-ups of characters' faces, so we focused on characters' non-verbal, non-facial behavior.

The first method is to modify a character's motion animation. Should the character walk quickly or slowly? Should the hands be clenched or relaxed? Should the shoulders be hunched? Should they walk with their head held high? We refer to this motion as gait, although technically, the term gait may not include specific hand and head movements. We have designed 5 different gait animations: neutral, angry, happy, sad, and afraid.

Figure 1: Screenshot from a scenario in the experiment.

The second method is to change how a character interacts with the game environment. We call these interactions emotional incidents. Our current emotional state influences the actions we take. For example, when happy, we may wave at friendly characters (see Figure 1). We want to ascribe emotions to the characters that affect how they interact and move about their environment so that players can perceive these emotions and ascribe motivations to the characters.

While we developed our animations (gaits, postures, gestures) to express emotions, they are not always emotion-specific. For example, we use the same kick animation for both the happy and angry emotional incidents. By combining multiple animations and changing the outcomes of the

Emotion	Response
Happy	Interact with prop
Angry	Move prop
Sad	Ignore prop
Afraid	Avoid prop

Table 1: Possible reactions for emotional prop.

Emotion	Response
Happy	Move to talk with another character
Angry	Direct path around obstacle
Sad	Ignore obstacles
Afraid	Buffer of space around obstacles

Table 2: Possible reactions for emotional path.

incidents, different emotions can be expressed. The posture/gait animations were largely based on the models developed by Roether *et al.* (2009). However, we also referred to various work done in psychology, such as the work of Lindquist and Barrett (2008), Mignault and Chaudhuri (2003) and Rossberg-Gempton and Poole (1993).

Emotional Incidents

We divided the emotional incidents into three categories: emotional props, emotional paths, and emotional characters.

Emotional Props An emotional prop may be a soccer ball, a rock, or a chair; i.e. any object in the scene with which a character can interact. The character's emotion will help determine how the character reacts. An example of character reactions based on emotion is shown in Table 1.

Emotional Paths Emotional paths refers to how a character travels between two points. For example, when maneuvering around a large obstacle, we believe angry characters will be likely to walk the most direct path; walking quite close to the obstacle. Characters who are afraid will try to keep a buffer of space between themselves and the obstacle, allowing them to see further ahead and prevent surprises. The happy and sad characters will be unlikely to have their paths influenced by the obstacle. However, happy characters will be more likely to delay following a path, or even change course, if they see a character or prop with which they want to interact. Table 2 provides a brief description of how a character's path may be influenced by each emotion.

Emotional Characters Finally, emotional characters refers to how a character interacts with other characters in the game world. Table 3 shows how each emotion may affect a character's interaction. For example, happy characters will greet and interact with other characters, while sad characters will tend to avoid them. Emotional characters is separate from emotional props because characters can have emotions themselves, which can complicate interactions. In addition, game characters often belong to groups of factions. Two happy characters who are from different hostile factions are not likely to show happiness towards each other.

Emotion	Response
Happy	Mimic positive gestures (wave if they wave)
Angry	Hostile acknowledgement
Sad	Ignore / no acknowledgement
Afraid	Avoid

Table 3: Possible reactions for emotional characters.

Figure 2: The four emotive gaits: Happy, Sad, Angry, Afraid.

These incident types will often be combined together. A scared character who encounters another character is likely to change its path to avoid walking too close to the other character. On the other hand, a happy character may change its path in order to interact with another character.

Experimental Design

In order to evaluate our approach, we designed an experiment to determine how accurately players can identify emotions based on non-verbal cues. We tested five gaits (neutral, happy, sad, angry, and afraid) and two emotional incidents. Figure 2 shows the four emotive gaits. In the first incident, the test character approaches a second character who is sitting on a bench and waving. In the second incident, the test character reacts to a child kicking a soccer ball.

The experiment contained thirteen different versions of the scenario as detailed in Table 4. Five of the scenarios contain a single gait and no incidents. These scenarios are similar to the experiment of Roether *et al.* (2009) and determine if participants can identify emotion through gait alone. The other eight scenarios contain the two incidents. Four of these pair an emotive gait with the appropriate emotional behavior. The other four use the neutral gait paired with an emotional behavior. Our goal was to determine if the incidents provided enough information for the participant to accurately identify the emotion without using an emotional gait.

Participants

The experiment had 44 participants; all undergraduate students taking a first year psychology class in the January 2012 semester. There were 36 females and 8 males with an average age of 20 (mode 19, range 18 - 35) and average year of study of 1.9 (mode 1, range 1 - 5). They received 2% credit in their psychology course for participating.

“Gait”	Happy	Sad	Angry	Afraid	Neutral	RSum	Recall
Happy	**18**	0	5	1	20	44	0.409
Sad	0	**22**	4	6	12	44	0.5
Angry	3	0	**27**	1	13	44	0.614
Afraid	1	0	1	**34**	8	44	0.773
PSum	22	22	37	42	53		
Precision	0.818	1	0.73	0.81		**0.821**	**0.574**

Table 5: “Gait” Confusion Matrix.

“Incident”	Happy	Sad	Angry	Afraid	Neutral	RSum	Recall
Happy	**34**	1	0	0	9	44	0.773
Sad	1	**36**	0	0	7	44	0.818
Angry	1	1	**30**	0	12	44	0.682
Afraid	1	9	5	**4**	25	44	0.091
PSum	37	47	35	4	53		
Precision	0.919	0.766	0.857	1		**0.846**	**0.591**

Table 6: “Incident” Confusion Matrix.

Version	Gait	Incidents	Emotion for Incidents
1	Neutral	No	-
2	Happy	No	-
3	Sad	No	-
4	Afraid	No	-
5	Angry	No	-
6	Neutral	Yes	Happy
7	Happy	Yes	Happy
8	Neutral	Yes	Sad
9	Sad	Yes	Sad
10	Neutral	Yes	Afraid
11	Afraid	Yes	Afraid
12	Neutral	Yes	Angry
13	Angry	Yes	Angry

Table 4: The thirteen scenarios.

Results

Confusion Matrices

To analyze our results, we created three confusion matrices (CMs): “Gait” only, “Incidents” only, and “Both”. CMs are a statistical tool (van Rijsbergen 1979) used to evaluate classifiers. They measure recall, precision and accuracy. In machine learning, a classifier is a machine learning technique. In our experiment the participants are the classifiers. The three CMs are presented in Tables 5, 6 and 7.

Each row represents a particular scene, and the row label (happy, sad, angry, afraid) indicates the emotion that the designer wanted the character to portray. The five columns (happy, sad, angry, afraid, or neutral) represent how participants classified a scene. RSum is the sum of the row; how many results for that scene we have. Recall is the ratio of how many participants correctly identified the scene divided by the RSum. The number of correctly identified emotions are on the diagonal, where the row and column labels match, and are shown in bold. PSum is the sum of a column; how many times participants used that label over all scenes. Finally, Precision is the ratio of how often participants used a label correctly divided by the PSum.

The overall precision and recall values for the table are in the bottom right corner. Overall recall is the total number of correctly identified emotions by all participants divided by the total number of participant/scene combinations (44 x 4). Overall precision is the total number of correctly identified emotions by all participants divided by the total number of participant/scene combinations not identified as neutral.

The CMs show that, generally, precision was quite high for each emotion in each matrix (the minimum precision was 0.73). This means that when a participant identified an emotion (non-neutral), there was a 73% chance that they identified the intended emotion. However, the recall values varied much more, even within individual CMs.

For example, in the “Incident” CM (Table 6), happy, sad, and angry all have a recall value ≥ 0.682, while afraid has a recall value of 0.091. The afraid recall value is the lowest in all three tables. These results were surprising. If you remove the afraid row, the average recall jumps to 0.758 from 0.591. Afraid is the only emotion that has a worse recall in the “Incident” table than in the “Gait” table. The participants depended on the gait to identify that a character was afraid. This result indicates that the best technique for conveying emotion may depend on the emotion being conveyed. Therefore, when other emotions (and states-of-mind such as confused) are introduced, similar tests should be done to determine the best technique for conveying them to the player.

T-Tests

As it is not possible to determine variance from confusion matrices, we used a statistical technique called bootstrapping to estimate variance to allow us to perform statistical analysis on the data.

Bootstrapping (Varian 2005; Services 2012) is a method for resampling data and produces ‘extra’ or ‘new’ data sets. It creates a new dataset by randomly choosing from the original set with replacement. In our case, our original set con-

"Both"	Happy	Sad	Angry	Afraid	Neutral	RSum	Recall
Happy	**39**	0	1	0	4	44	0.886
Sad	1	**36**	4	1	2	44	0.818
Angry	2	0	**41**	1	0	44	0.932
Afraid	0	0	1	**38**	5	44	0.864
PSum	42	36	47	40	11		
Precision	0.929	1	0.872	0.95		**0.933**	**0.875**

Table 7: "Both" Confusion Matrix.

	Recall	Precision
Overall		
"Both" vs "Incident"	**1.245×10^{-14}**	**4.333×10^{-8}**
"Both" vs "Gait"	**2.377×10^{-13}**	**2.342×10^{-9}**
"Incident" vs "Gait"	0.474	0.301
Happy		
"Both" vs "Incident"	**1.025×10^{-4}**	0.417
"Both" vs "Gait"	**1.211×10^{-13}**	0.0170
"Incident" vs "Gait"	**6.435×10^{-11}**	0.0278
Sad		
"Both" vs "Incident"	0.124	**3.047×10^{-9}**
"Both" vs "Gait"	**6.052×10^{-8}**	*both rows are all 1s
"Incident" vs "Gait"	**7.052×10^{-8}**	**3.047×10^{-9}**
Angry		
"Both" vs "Incident"	**8.424×10^{-14}**	0.125
"Both" vs "Gait"	**3.906×10^{-11}**	**4.031×10^{-7}**
"Incident" vs "Gait"	0.243	**4.338×10^{-6}**
Afraid		
"Both" vs "Incident"	**2.918×10^{-16}**	**3.449×10^{-4}**
"Both" vs "Gait"	**5.980×10^{-4}**	**1.736×10^{-6}**
"Incident" vs "Gait"	**1.259×10^{-17}**	**5.688×10^{-7}**

Table 8: T-Test Results. (Bold = 99% confidence)

sisted of 44 rows, each row representing how an individual participant scored the thirteen scenes. By using bootstrapping, we created 10 'new' data sets of 44 rows. Each row in a 'new' dataset was a row from the original dataset. Because bootstrapping produces new data by choosing with replacement, a new data set may contain four copies of row 22 from the original set. If we did not use replacement, each new data set would be identical to the original.

For each new data set we produced, we created confusion matrices similar to the ones in Tables 5, 6 and 7. Using the accuracy and precision values from our 11 datasets (10 new + original), we were able to run a series of T-Tests to test for 99% confidence. Table 8 shows the results of the T-Tests.

From the raw data, we expected the "Both" results to be significantly better than either "Gait" or "Incidents" on their own, which the T-Test confirmed. We were not surprised that results for "Incident" versus "Gait" were insignificant.

The results for individual emotions were interesting, as shown in Table 9. Afraid, the only emotion that produced statistically significant results in all six tests, gives a different ranking for recall than precision. For afraid, "Gait" produced higher recall than "Incident", while "Incident" produced higher precision than "Gait". Happy, the only other emotion to get statistical significance in all three recall values, gives a ranking of "Both" > "Incident" > "Gait".

	Prefer Recall	Prefer Precision
Overall:	Both >> Gait ≈ Incident	
Happy:	Both > Incident > Gait	Both ≈ Incident ≈ Gait
Sad:	Both ≈ Incident > Gait	Both ≈ Gait > Incident
Angry:	Both > Incident ≈ Gait	Both ≈ Incident > Gait
Afraid:	Both > Gait > Incident	Incident > Both > Gait

Table 9: Ranking of three methods

The results suggest that while overall it is much better to use "Both" than either "Gait" or "Incident" on their own, for some emotions, it is possible to get similar results using only "Gait" or "Incidents".

Which method(s) to use?

It is not immediately clear which method one should use so that players can most accurately identify the emotions of background characters. The best method will depend on what matters more in the game: precision or recall. The relative importance of precision and recall may be highly dependent on the type of game being created. Table 9 shows the ranking of the three methods depending on whether one values recall or precision.

From the table, it appears that in order to have a high recall value, it is generally necessary to use both the gait and incidents. Only the sad emotion had "Incident" scores as high as "Both" scores. However, for precision measures, the individual emotions produced equally good results using either "Incident" or "Gait" on their own or using "Both".

"Gait", "Incident" and "Both" present different costs to the game designers. "Gait" will require a minimum of four gait animations per character type to represent each of the four emotions. If characters perform one of the four gait animations, game players may learn to ignore the emotion, so a neutral gait is probably necessary to maintain the impact of the emotional gaits. On the other hand, using incidents requires a larger set of animations in order to perform the various incidents, as well as identifying game objects for use in incidents. However, the incidents do not happen all the time, which creates a more unpredictable atmosphere. Finally, using both the gait and the incidents together requires creating both sets of animations. Overall, the combined method produced the strongest results, but the cost of implementation suggests that game designers should focus on the best technique for each emotion.

Entertainment Games Entertainment games could use emotion to enhance player experience. However, adding emotion is currently seen as optional. Therefore, we believe that most entertainment game designers will be more concerned about precision than recall. That is, a designer will want to know that if a player ascribes an emotion to a character, then it was the intended emotion. Since current games rarely provide emotional cues, most designers will probably be less concerned about a player missing a cue.

Serious Games On the other hand, for serious or training games, we believe the opposite is true: recall matters more than precision. A high recall value means that players more often recognize that characters are emotional. With high recall, if a character is supposed to be angry, the player is more likely to notice. If an emotion is being displayed in a serious game, it is usually for a very specific reason and it is important for the player to identify (recall) it.

However, at the expense of a high recall value, players may ascribe an emotion that the designer was not trying to display. For example, a player identifying a Neutral character as Happy is the cost for high recall, with low precision.

Anecdotes

Anecdotal accounts from the study participants indicate that each participant focused on very different characteristics, and identified different patterns and observations as important for classification. For example, in the happy interactions, the character kicks the soccer ball back to the child, while the angry character kicks the soccer ball away from the child. Some participants thought this was emotionally relevant, while others thought the difference had to do with the character not being accurate when kicking the ball.

A second example is based on the arm and hand positioning of the character. An angry character walks with clenched fists, while the character who is afraid has their hands slightly in front of them, with splayed fingers. While some participants used that information to identify all four emotions, other participants ignored or did not notice this clue. Figure 2 shows the different hand and arm positions.

When the angry character passes the man on the bench, she turns to face him and speeds up. Some participants remarked that the "angry character glares at the man", which is notable, as the player only sees the back of the character, and cannot see the character's face. This strongly suggests that participants were picking up on the emotion, and were mentally ascribing extra non-verbal behaviors.

Conclusion

Providing characters with identifiable emotions through their non-verbal behavior will benefit games created strictly for entertainment and those designed to teach or train the player. Identifiable emotions will allow players to feel more connected to the game world, make the game world more believable, and support deeper story lines. The results of our study show that the combination of both gait and incidents generally provides the best results. However, for some emotions, gait or incidents alone can provide equally good results. The next step in our research is to implement a system that will allow game designers to easily assign emotion to background characters without having to hand script them.

Acknowledgements

We thank GRAND, NSERC, Autodesk, and Alberta Innovates Technology Futures for funding this work and Jamie Schmitt for creating the models and animations. Finally, thanks to all members of the Believe team at the University of Alberta.

References

Bethesda. 2011. Skyrim. Website. http://www.elderscrolls.com/skyrim/.

BioWare. 2012. Mass effect 3. Website. http://masseffect.bioware.com/.

Cao, Y.; Tien, W. C.; Faloutsos, P.; and Pighin, F. 2005. Expressive speech-driven facial animation. *ACM Trans. Graph.* 24:1283–1302.

Desai, N.; Szafron, D.; Sayegh, L.; Greiner, R.; and Turecki, G. 2011. Challenges and problems of prototyping a serious game. In *Annual GRAND Conference*. Vancouver, Canada.

Lindquist, K. A., and Barrett, L. F. 2008. Constructing Emotion The Experience of Fear as a Conceptual Act. *Psychological Science* 19(9):898–903+.

Mignault, A., and Chaudhuri, A. 2003. The Many Faces of a Neutral Face: Head Tilt and Perception of Dominance and Emotion. *Journal of Nonverbal Behavior* 27(2):111–132.

Ortony, A., and Turner, T. 1990. What's basic about basic emotions? *Psychological Review* 97:315–331.

Perlin, K. 2011. Polly's world. Website. http://mrl.nyu.edu/ perlin/experiments/polly/.

Pixar. 1986. Luxor jr. Website. http://www.pixar.com/shorts/ljr/.

Roether, C.; Omlor, L.; Christensen, A.; and Giese, M. A. 2009. Critical features for the perception of emotion from gait. *Journal of Vision* 9(6):1–32. reviewed.

Rossberg-Gempton, I., and Poole, G. D. 1993. The effect of open and closed postures on pleasant and unpleasant emotions. *The Arts in Psychotherapy* 20(1):75 – 82. Special Issue Research in the Creative Arts Therapies.

Seif El-Nasr, M., and Wei, H. 2008. Exploring non-verbal behavior models for believable characters. In *Proceedings of the 1st Joint International Conference on Interactive Digital Storytelling: Interactive Storytelling*, ICIDS '08, 71–82. Berlin, Heidelberg: Springer-Verlag.

Services, U. A. T. 2012. R library: Introduction to bootstrapping.

van Rijsbergen, C. 1979. Information retrieval. Website. http://www.dcs.gla.ac.uk/Keith/Preface.html.

Varian, H. 2005. Boostrap tutorial. *The Mathematica Journal* 9(4):768–775.

Wallbott, H. G., and Scherer, K. R. 1986. Cues and channels in emotion recognition. *Journal of Personality and Social Psychology* 51(4):690–699.