Learning to Influence Emotional Responses for Interactive Storytelling

David L. Roberts and **Harikrishna Narayanan** and **Charles L. Isbell**
School of Interactive Computing
College of Computing
Georgia Institute of Technology
robertsd@cc.gatech.edu, harikrishna@gatech.edu, isbell@cc.gatech.edu

Abstract

We present an architecture for interactive storytelling. The system interleaves pre-authored text with pre-selected videos to generate a story. Between iterations, the player is given an opportunity to answer questions that help to drive the narrative. Videos are used to have an effect on the emotional response of the players. The system is capable of performing modeling of both videos and players to better adapt the narrative progression in response to the player's answers to questions.

The system is designed to serve two purposes: 1) to label natural language utterances and passages for use in a text classification system; and 2) to serve as a test environment for computational models of influence and persuasion. We motivate the approach we have taken by describing research efforts on classifying emotion in natural language and on the use of influence to affect decision making. The architecture and prototype system is described in detail. A summary of human-subject experiments planned are included as well.

Introduction

In this paper we present an architecture for authoring and interactively telling stories using a combination of videos and text. The system is designed to serve two purposes: 1) to create and observe emotions in humans while generating a labeled data set for a machine learning/natural language processing system; and 2) to test the effect of computational models of influence based on theories from social psychology and behavioral economics.

Stories play a significant role in the way we communicate. On a personal level, we use stories to inform others of important events in our lives and as a window into our emotional state. We are a culture of story tellers. In fact, we are so practiced at storytelling that we can adjust our approach in reaction to our audience. We can read reactions like eye contact, facial ticks or positions, body posture, and response to questions as we tell our stories. All of these signs that we interpret enable us to adapt the details of our story to try to engage the listener, establish a rapport with them, and ideally ensure they experience the exact reaction to our story that we intend for them to experience.

In addition, artists and authors (and an increasing number of game designers) use story as a tool to create an alternate reality for people. A skilled storyteller has the ability to cause the present to fade and construct a fictional world that the listener(s) become immersed in. In doing so, the storyteller can be very influential to the listener, causing a range of emotions—some of which may extend long after the telling of the story has completed.

In traditional storytelling mediums such as print and film, the listener is a mostly passive participant in the experience. With the exception of the minor adjustments they can make by using their own imagination, listeners have no affect on the outcome or style in which the story is told. In recent years, computer hardware technology has reached a point that enables real-time 3d graphics, complex virtual characters, sophisticated AI algorithms, and other approaches to be used in the design of *interactive* virtual experiences where the listener takes on a more active role in the story telling process. This active role, in effect, changes the passive listener into an active *player*.

It is important to note that when a *listener* becomes a *player* a set of challenges arise that can potentially be significant hurdles for the story author. First, the author must take steps to create "setting" and "characters", which we refer to as the *model*. Given the choice of model, the author must specify what, out of everything consistent with the model, is consistent with their aesthetic or artistic *goals* for the story. For example, if the author is creating an interactive mystery drama, the model would consist of the location (perhaps a town), structures in the environment (perhaps a few buildings, streets, cars, etc.) and characters that populate the environment (some that are important the plot and some that may not be). The choice of model is not particularly specific and in practice the author's goals for the experience represent a subset of possible interactions that are consistent with the model. The ability for the player to interact, or more specifically exercise their sense of *self-agency*, allows for goal-driven exploration on the part of the player. In the case of live storytelling this is tantamount to allowing listeners to ask questions and direct the storyteller. The degree to which the player's self-agency can lead them to parts of the space that are not consistent with the author's goals is the degree to which *threats* to the narrative exist.

Threats in an interactive storytelling environment necessi-

tate the use of what has been referred to as a *drama manager* (DM) in the literature (Laurel 1986). A drama manager is an agent of the author that is tasked with shaping the player's experience. In part, the drama manager functions like a human storyteller would if they were adapting their story based on their observations of the listeners. Throughout the remainder of this paper we will present an interactive storytelling architecture designed to study the effect of influence, emotions, and the adaption of stories based on emotions.

Why Emotions?

Emotions are vital to what makes us human. They enhance and augment our ability to experience our environment and our relationships with others. A well told story can have a profound effect on us. It can shape our perspective on important issues or influence our opinion of people. The most effective storytellers create an emotional connection with their audience. They do so by describing characters or events in their stories that listeners can identify with. In part, listeners latch onto the emotional response that characters have to story events or to other characters. There are two flavors of research related to emotion in a virtual setting: generation and recognition.

There has been a lot of work on modeling emotion for virtual characters in the intelligent virtual agent's community. When embodied virtual characters are used to tell a story, it is important that they appear believable. Subtle verbal and non-verbal behaviors in virtual characters can convey deep emotion. Thus, for embodied characters generating emotion is a critical task (Gratch and Marsella 2001). A character's emotional state has an effect on their decision-making, actions, memory, attention, voluntary muscles, etc. (Berkowitz 2000). Because emotion is portrayed by such a wide array of behaviors, we can learn a lot from the actions people take without hearing their words.

On the other side of the spectrum, and more closely related to this paper, is work on recognizing emotion in humans. There are a variety of approaches to this task using a number of modalities. Some approaches make use of techniques from computer vision that are based on non-verbal gestures such as gaze aversion (Morency, Christoudias, and Darrell 2006) and head pose (Stiefelhagen 2002). In more recent work, the vision algorithms are augmented with conversational context to improve the recognition accuracy (Morency et al. 2007). Using context that is available by analyzing natural language (either recognized or generated) or simply world state information can aide significantly in the interpretation of recognized gestures.

In addition to recognition of non-verbal behaviors, recognizing prosody is a key element to interpreting emotion. Prosody is the quality of speech as observed by pitch, intonation, duration, rhythm, volume, etc. Prosodic cues can be equally informative when classifying utterances according to emotional state. Dellaert, Polzin, and Waibel present an approach based on smoothing pitch in voice recordings of utterances (1996). Interactive voice response systems (such as automated telephone call centers) have been used to study speech classification based on anger or neutrality. A number of machine learning methods such as artificial neural networks, support vector machines, and decision trees have been applied to that domain (Yacoub et al. 2003); however, it has been shown that prosodic cues alone do not yield accurate results in real world scenarios (Huber et al. 2000). The additional use of context similar to Morency et al. (2007) has been demonstrated to be warranted in these situations.

Lastly, there have been a number of research efforts to recognize emotion using physiological measurements such as heart rate, skin conductance, and facial electromyogram, which are summarized by Lisetti and Nasoz (2004). Those studies were conducted using a "noninvasive" wearable computer. They achieved classification accuracy in the $80\% - 90\%$ range for most of the six emotions they were attempting to classify: sadness, anger, surprise, fear, frustration, and amusement. A relatively high success rate such as this is more than adequate for a majority of interactive entertainment settings.

There have been some efforts to analyze emotion from text (as apposed to speech). Most characterize emotion with low granularity into five categories such as: high arousal, negative; low arousal, negative; neutral; low arousal, positive; and high arousal, positive (Osherenko 2008). These methods tend to rely on an "affect dictionary" that provides relationships between words and emotions (Read 2004). There has been some effort to learn an affect dictionary from general dictionaries, but the results of experiments indicate that domain-specific affect dictionaries tailored to the corpora used results in higher classification accuracy (Osherenko and Andr 2007).

To avoid the problems of affect dictionaries, some approaches have leveraged commonsense reasoning based on ontologies of real world facts combined with lexical models (Liu, Lieberman, and Selker 2003). Neviarouskaya, Prendinger, and Ishizuka present an approach also based on ontological knowledge for characterizing one of nine emotions in "journal style" blog posts (2007). To measure success, the authors have three human participants hand classify each sentence in the corpus into one of the nine emotions they are trying to identify.

The idea of using humans playing a game to label data for machine use is similar to von Ahn's notion of *human computation* (2006). von Ahn, along with his collaborators, have created web games such as Peekaboom (von Ahn, Liu, and Blum 2006) and Verbosity (von Ahn, Kedia, and Blum 2006) to get humans to label images based on their contents and to acquire common-sense knowledge—two things machines are not good at. Similarly, we hope to leverage interactive story participants to perform similar tasks. We will discuss in more detail below the ways in which our system can help to create labeled corpora for these types of affect/emotion recognition experiments without the need for explicit human coders.

Why Influence?

As described above, the task of effectively balancing between an author's goals for an interactive story and the player's goals is very important. This task amounts to the minimization of potential threats. There have been numerous representations of story and associated algorithms for

solving this problem collectively referred to as experience or drama managers (Roberts and Isbell 2008; Mateas 1999). A majority of the work on drama management has been on abstract representations of story (with a few exceptions) and results presented are of experiments conducted in simulation.

The use of simulation to evaluate drama mangers has led to some criticism of the result's applicability to real-world interactive experiences. For example, the Declarative Optimization-based Drama Manager (DODM) first proposed by Bates (1992) and studied more extensively later by Weyhrauch (1997) uses drama manager actions to affect the plot progression in a story. These actions are modeled as abstract operators on significant plot events such as *cause*, *deny*, or *hint*. A variety of algorithms have been used to solve a DODM instance including search (Weyhrauch 1997), Reinforcement Learning (Nelson et al. 2006), and a variant of Markov decision processes known as TTD-MDPs (Roberts et al. 2006; 2007). Recently these results have come under scrutiny for relying heavily on the drama manger's cause action and therefore being overly manipulative and reducing the player's choices—in effect limiting their self-agency (Nelson and Mateas 2008).

While Nelson and Mateas target their critique on the TTD-MDP-based DODM solution technique, it is likely the case that their criticism applies to many of the approaches to drama management that rely on proactive actions. They argue that one way to address the issue is to change the objective of the drama manager to explicitly include a representation of player choice (and therefore self-agency) and present the results of a brief simulation experiment to support this claim.[1]

Another approach to addressing the issue of drama managers being overly manipulative (regardless of the particular approach) is to look more carefully at the implementation of the drama manager's actions. Specifically the types of actions that cause an event in the story world. Roberts et al. (2008) propose the use of computational models based on the theories of *influence* and *persuasion* from social psychology (Cialdini 1998) and behavioral economics (Ariely 2008) as a tool for automatically refining drama manager actions. They argue that a significant feature of actions based on those theories is the strict preservation of the player's sense of self-agency. That argument is summarized here.

Because interactive experiences are marked by a strong social context, in order to fully engage in the management of these experiences it is important to move beyond simple physical manipulation of the environment. The goal in developing computational models of influence is to: 1) benefit authors by providing tools designed to influence players to buy into the adoption of goals consistent with the author's; 2) reduce the burden on authors by enabling them to specify goals abstractly, relying on the principles of influence and

persuasion to bridge the gap to a concrete implementation in the virtual environment; and 3) accomplish (1) and (2) without the player perceiving any decrease in (and preferably an increase in) self-agency.

It should be noted that the traditional approach of physical manipulation of the environment can put players in situations that may change their mental or emotional state provided the experiential content has been appropriately authored. It is relatively straightforward to author for the transfer of knowledge from an NPC to the player; however, simply imparting knowledge to the player is not sufficient for increasing the likelihood that they will choose to adapt a specific goal. Further, using computational models of influence the drama manager will be able to decide how to change the player's mental or emotional state without using detailed pre-authored content. The main benefits of this are twofold: 1) With the notable exception of the beat-based drama manager (Mateas and Stern 2005), no other approach to drama management has been designed to influence a player's goals which means new doors to creating narrative experiences may open; and 2) The drama manager can try to align the player's goals with the author's goals without them realizing it is doing so—thus preserving their perception of self-agency while increasing the affordance for authorial control of the narrative experience.

There are six major principles of influence to focus on. These principles have been identified by years of research in the field of social psychology and behavioral economics and are frequently employed as sales tactics by savvy marketers. They are:

- **Reciprocation**: give and take; when someone does something for us we feel obligated to return in kind.

- **Consistency**: we have a near obsessive desire to be (and appear) consistent with what we have already done or said.

- **Social Proof**: we look to others like us to determine the appropriate action to take.

- **Liking**: the more we like someone, the more likely we are to abide by her requests.

- **Authority**: we have a deep-seated sense of duty to authority.

- **Scarcity**: something that, on its own merits, holds little appeal to us will become decidedly more enticing if it will soon become unavailable to us.

The principles provide the foundation for understanding how to create the powerful tools used on a daily basis by sales professionals. Used properly by themselves or in combination with each other these principles can greatly increase the likelihood of someone complying with a request. The fact that these principles can only increase the likelihood of compliance is an important feature of an approach to drama management that uses these ideas. The player can always decide not to do what the drama manager is trying to get her to do. Thus, while the careful application of these principles of influence can greatly increase the chances of a player choosing to act in a manner the author prescribes, the affordance for self-agency is strictly preserved.

---

[1]One piece of the argument not discussed by Nelson and Mateas is the availability of non-manipulative actions such as hints to the drama manager. It is unclear if their results can be duplicated when the drama manager is given a sufficient number of appropriately modeled hints.
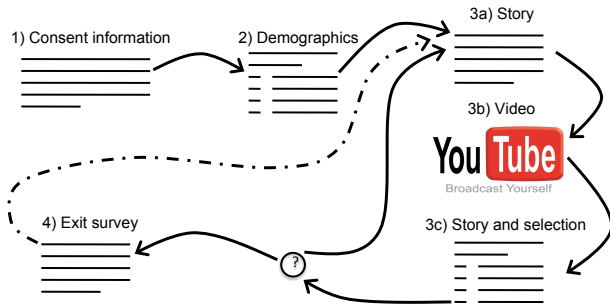
Figure 1: A schematic overview of the YouTube narrative system.

Figure 2: A screenshot of our prototype system during the question asking phase.

In order to use these principles effectively for interactive experiences, it is only necessary for the drama manager to hit upon the triggers that cause humans to behave in a predictable manner. Perhaps the most powerful thing about using the principles of influence to the DM's advantage is that to do so requires minimal effort. Therefore, a player willingly complying with the DM's wishes will tend to see her actions as a result of either her own choices or of natural forces rather than the influence of an exploiter.

Note that each principle can potentially be used in more than one way. For example, scarcity can be used to entice a player to obtain a particular object or to convince her that certain information she has obtained (or should obtain) is more important. Liking can actually be employed as a function of friendship, reputation, or physical attractiveness. Authority can be merely a title like Dr. or can be a function of celebrity.

Architecture

The design of our narrative system was created specifically to meet two goals: 1) to tell stories interactively that will cause participants to experience specifically targeted emotions and leverage those emotions to create a labeled data set for use in classifying free text based on the emotional state of the writer; and 2) to provide a test platform for analyzing the effectiveness of using computational models constructed upon the theories of influence and persuasion as the basis for drama manager actions that guide the player's experience in the interactive story.

To achieve those goals, our system uses a web-based interface that displays a sequence of authored text and videos that are interspersed with specific decision points for the player. The videos have been obtained from YouTube[2], a free online repository for streaming video. The architecture is presented in Figure 1. A player's experience will advance according to the following procedure (depicted in the figure):

1) **Consent information**: Upon arriving at the landing page, the player is presented with the study's consent information and asked to acknowledge it.

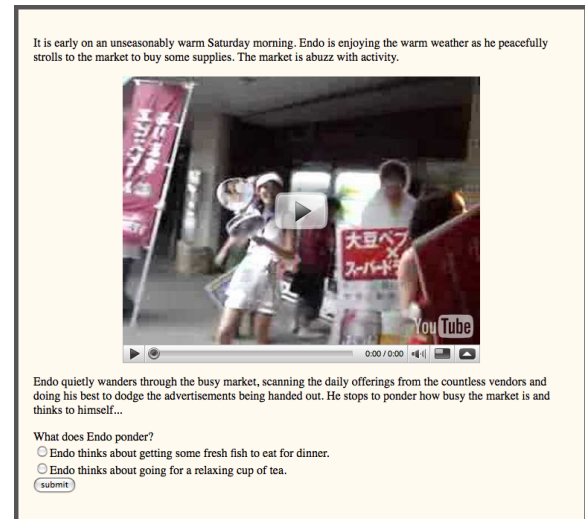2) **Demographics**: The player is asked to answer a set of basic demographic questions which will enable us to compare the groups of people in different treatments. Upon completion, the player is presented with brief instructions for continuing.

3a) **Story**: The player sees author-provided text that describes the beginning of a story event.

3b) **Video**: The player is presented with a video from YouTube presenting information supplemental to the story. Sights, sounds, and action can be conveyed much more efficiently. The video does not start playing automatically; it is completely optional for the player to view it.

3c) **Story and selection**: After the video, the player is presented with a short bit of text and a multiple choice question. The question solicits a decision from the player that will drive the narrative. At this point, the system cycles back to Step (3a) and displays another text-video-text set or, once the story has reached a conclusion, moves on to the final step.

4) **Exit survey**: The player is presented with a set of exit questions and, depending on the study, is asked to summarize the story they just experienced.

All of the text, question-response sets, and videos are pre-authored—not generated. We have opted to obtain the videos for our system from YouTube to ease the authoring process. Information about each video is stored in a database including: a unique identifier; the html code to embed the video in a webpage; any meta-data the author associates with it; and information on how often it is played. Additionally, the database is populated with the questions the player will answer, the set of potential responses, and the text authored for transitioning between events. The story structure is encoded in the database as well.

---

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

A majority of the responses to questions are categorical in nature. We have questions that are both explicit and subtle in the elicitation of emotional response categories. For example, to be explicit we may ask the player: "How do you feel about the actions of Thomas in the last video?" To be more subtle, we may ask the player: "What would you like to do in response to Thomas' actions in the last video?" In the former case, the set of responses will correspond to emotional labels such as "satisfied" or "disgusted." In the latter case, the actions availabe as responses to the player will be at the level of *discourse acts* such as "punch Thomas" or "hug Thomas." These discourse acts can be correlated to emotional responses as well. In this case, "punch" would correspond to anger or disgust whereas "hug" would correspond to empathy or happiness. The choice of discourse acts will determine the narrative progress as they correspond directly to player actions. Note that depending on the particular experiment we are running, both types of questions need not be asked of the player; however, in all cases a subtle question that leads to discourse acts will be presented to the player in order to determine how to continue the story.

The Use of Videos

As we will discuss further in the concluding section of this paper, we are not just interested in the creation of interactive stories, but also the authoring process. As skilled as we are at telling stories in general, very few of us are capable writers. We can retell stories that we have experienced or heard before, but putting those stories down on paper for others to read is something very few of us can do effectively. We believe that the limitations many of us have in disseminating our stories through text similarly applies to an interactive setting. Thus, an authoring paradigm that alleviates most of the burden associated with creating a written narrative is desirable. Further, with the increasing availability of personal video recording devices (even many cell phones can record video clips now) many people have vast collections of small snippets of their life's story. Weaving those videos together effectively is a much easier task than telling the entire story from scratch. Thus, one benefit of using videos for our interactive storytelling system is that we believe many more people will be able to use it tell stories.

Research into the creation of virtual humans is an ongoing area that, while having made much progress in recent years, still is far from the effectiveness of photo-realism one gets from actual humans in video. Because an emotional connection with characters is one way in which listeners (or players) can really engage in a storytelling event, using characters that may not appear realistic can adversely affect the quality of experience. Further, because virtual human technology is still very new, the engineering effort to create such a character is massive in comparison to that of creating a corpus of videos. Creating a corpus of videos is something that an average computer user can likely do effectively while building a virtual human is better left for technical experts well-versed in state of the art technology and algorithms.

Lastly, the emotionally persuasive power of video is well documented. As early as 1962, researchers were using video in lab experiments to study emotions such as stress (Lazarus

et al. 1962). More recently, film clips have been used to study a range of emotions. To examine the effectiveness of film at causing emotion, psychology researchers have compared "self-reports" among study participants to determine the consistency of emotional responses across subjects to particular video clips (McHugo, Smith, and Lanzetta 1982). Philippot has specifically studied the ability of videos to induce differentiated emotional states in lab experiments (1993). Others have studied the process by which an effective corpus of videos can be constructed specifically for use in emotion experiments involving human subjects (Gross and Levenson 1995). Finally, most recently, the effectiveness of self-reports for a constructed corpus of emotionally persuasive videos has been studied using classification of physiological responses such as heart rate, skin temperature, and pupil dilation (Lisetti and Nasoz 2004).

Thus, the use of videos in our system is a strategic choice that serves three important purposes: 1) they provide a simple authoring paradigm; 2) we can avoid the creation of virtual human characters; and 3) videos are emotionally persuasive. These three benefits to using videos in our storytelling system will help to enable a range of experiments on player experience, collection of labeled emotional utterances, and enable us to examine how unskilled authors can take advantage of the tools we provide.

The Use of Influence

Note that while videos are emotionally "persuasive," we are in fact interested in a different type of persuasion as well. A separate, but equal interest that drives the development of our system is to study the effect of computational models of influence and persuasion as tools for the automatic refinement of drama manager actions. We are specifically interested in the type of influence often used in marketing that can affect consumer behavior.

There are two questions associated with the use of influence in an interactive setting: 1) What specific tool of influence is appropriate to use given the current goal? and 2) What is the best way to refine the tool given the current state of the story? Thus, the drama manager's job is to select a goal for the story. Given that goal, a range of action "types" are available for the system to take. These actions types represent a model based on a particular theory of influence. Given an action type, the system then must choose how to implement (or refine) an instance of that action type given the current state of the story.

The YouTube storytelling architecture described in this paper enables experiments to be run that will test the effectiveness of solutions to both influence selection questions. In the concluding section of this paper we will describe the human-subject experiments we have planned in more detail. In short, we are developing specific models of a number of methods of influence. Each of these methods of influence will be tested against a baseline to determine to what degree their use can affect the action selection by players.

It should be noted that the models of influence for virtual characters have been implemented before. PsychSim, a social simulation system developed by Marsella, Pynadath, and Read, uses theory of mind for virtual agents in a partially

(a) Clear

(b) Ambiguous

Figure 3: The dogs pictured here are Shiba Inus, a breed known for hectic play style. Few would confuse the action on the left, but some might assume the shot on the right is of dogs fighting when in fact it is play. The difference in interpretation is likely manifest itself in different emotional responses. Photos courtesy of BW Pet Photography (http://bwpetphotography.blogspot.com).

observable Markov decision process framework to model the beliefs and goals of other agents in the simulation (2004). To maintain an agent's beliefs about the world and other agents, rules based on Cialdini's methods of influence are used to help explain the observed actions of agents and as a heuristic to select from alternative explanations of those observations. PsychSim has also been used in the Thespian interactive narrative system (Si, Marsella, and Pynadath 2005; 2006). The models of influence that we will test using our YouTube storytelling system are different than those of PsychSim. PsychSim's models are designed for agent to agent interactions and are used to describe behavior observed in other agents. The models we will test using our system are designed for agent to human interaction and are used to refine agent behaviors that will affect, rather than explain, human behavior.

Player Modeling and Adaption

One key feature of an interactive storytelling system is its ability to adapt to the user. The YouTube storytelling architecture is no exception to this rule. The author's goals in this setting are represented as a set of emotional changes over time. The task for the system is to present videos and text in sequence to induce those affect changes. Our system will use a Declarative Optimization-based Drama Manager (DODM) described above. A detailed description of the DODM approach to managing the experience is beyond the scope of this paper.

For the DODM manager to be effective, it is necessary for the system to model the player in a way that allows it to reason about the author's goals and the player's behavior in relation to those goals. Since the author's goals are represented as emotional change over time, a model the player's emotional reactions is vital to proper function. Therefore, the YouTube storytelling system maintains a cumulative model of player responses to videos over time.

Note, it is possible for players to interpret certain videos in different ways, and therefore experience different emotional reactions. For example, since we cannot show a video in this paper, consider the still photos of dogs in Figure 3. The Shiba Inus in those photos are a breed known for their apparently aggressive play style. In Figure 3(a) the lack of action and calmness is likely to elicit a positive, or happy, response from viewers; However, if the viewer has a fear of dogs, they may respond very differently. Further, in Figure 3(b) the dogs are at play. The commotion conveyed in the image and the visible teethe would be construed by most as a sign of fighting and generally cause a concerned or upset reaction. On the other hand, those familiar with the breed and its play style might recognize it as playing and react happy.

The player's reactions will be tracked over time in two ways. First, for each video and bit of authored text in the system, a cumulative reaction count is maintained to model the average emotional response of players. Assume we have k emotions $e_1, \ldots, e_k$ and l actions $a_1, \ldots, a_l$. Let $c(e_i, v_j)$ be the number of times emotion e_i is reported in association with video v_j in response to an explicit question, and $c(a_i, v_j)$ be the number of times action a_i is taken in response to a subtle question about video v_j. Note that a player's actions need not correspond one-to-one with a particular emotion. For example, a "punch" action could indicate fear or anger. Thus, let $f(a_i, e_j) \in [0, 1]$, $\sum_{m=1}^{k} f(a_i, e_m) = 1.0$ be an author defined function that represents the emotional responses associated with action a_i. This function captures the multiple-interpretation effect that can occur in response to videos. Then, assuming video v_i has been viewed n times, the model of video v_i is given by the vector:

$$\vec{V_i} = \left[\ldots, \frac{c(e_j, v_i) + \sum_{m=1}^{l} f(a_m, e_j) \cdot c(a_m, v_i)}{2n}, \ldots \right]$$

This vector represents the average emotional responses for the video across all player and all episodes. By itself, this

does not provide a model the player, but a model of how the average player will respond to the video.

To construct the player model, a comparison of the player's particular emotional responses to the average across all players' responses is necessary. Thus, a player's self-reported emotional response and action define a player response vector for a video v_i:

$$\vec{P}_i = \left[\cdots, \frac{s(e_j, v_i) + \sum_{m=1}^{l} f(a_m, e_j) \cdot s(a_m, v_i)}{2}, \cdots \right]$$

where s is a selection function that evaluates to 1.0 when the player has self-reported the emotion or taken the action passed as the first argument in response to the video passed as the second argument and evaluates to 0.0 otherwise. $\vec{P}_i$ represents the player's emotional response to video v_i and can easily be compared to the model $\vec{V}_i$ for the video.

Simple vector subtraction provides a dissimilarity measure $\vec{D}_i = \vec{V}_i - \vec{P}_i$ of the player model from the video model. Thus, when selecting a video to present to the player, the system can reason about how the player's emotional reactions are likely to be different from the expected. If the system has a goal of creating an emotional response that corresponds to emotion e_j it can examine videos the player has seen already that, on average, cause emotion e_j and determine if the player reacts similarly to the general population of players. If so, it will show that video. If not, it will search through the dissimilarity vectors it has computed thus far to determine if there is another type of video in the system that may have better success causing emotion e_j.

Concluding Thoughts

In this paper, we have described our YouTube storytelling system. The system uses a combination of pre-authored text, question-response pairs, and a corpus of videos from YouTube to tell stories. The system will enable two orthogonal directions of future research—analysing free text to classify the emotional state of the writer and using computational models of influence to affect the choice of actions by players.

We have a functioning prototype of the system that includes the majority of the features described in this paper. We plan to begin evaluating the system in the near future. With only one minor exception, all of the evaluation of the YouTube interactive storytelling system will involve human subjects. First, we plan to run "baseline" studies with a fixed story structure. This baseline study will serve three purposes. First, we will collect data to build the video models described in the previous section. Once we have meaningful models, we can begin to perform the adaption that is characteristic of interactive storytelling system. Second, we will have a small set of sentences that are labeled by the emotional state of the writer. Eventually, this data set will be used to train a classifier that will work on natural language input into the YouTube narrative system in place of the multiple choice questions we are currently employing. Third, it will provide a data set free from the type of marketing influence we will test in subsequent studies.

The overlap between evaluation of our work on emotion and on influence will end after the first study. The second study we plan to conduct will include a DODM drama manager, player modeling and adaption, and will have branching story structure. This study will be conducted specifically to characterize the ability of the drama manager to cause emotional change over time in the player. It will further serve to gather more data for our natural language classifier.

Next, we will perform a series of experiments to test various influence techniques (or combinations of techniques). These influence techniques will be constructed as actions available to the drama manager that will change the text (and potentially questions depending on the particular technique) that is presented to the player in a principled manner. The intended result of these experiments is to demonstrate a measurable and statistically significant change in the frequency of actions taken by the player given the same fixed story structure presented to the players in the baseline study.

Lastly, we plan to study how author's can leverage the tools we have developed for this system. We plan to conduct a study where participants are asked to compose text and question/response sets as well as select videos for inclusion in an interactive story they author. We are interested in getting author's impressions of how useful it is to have influence as a tool available to them. The authors participating in the study will be given access to "influence tools" they can associate with explicit goals they have for their interactive story. These influence tools will be directives to the system about a technique it should use to realize the author's goal, the details of which will be generated automatically.

References

Ariely, D. 2008. *Predictably Irrational: The Hidden Forces That Shape Our Decisions.* Harper Collins.

Bates, J. 1992. Virtual reality, art, and entertainment. *Presence: The Journal of Teleoperators and Virtual Environments* 1(1):133–138.

Berkowitz, L. 2000. *Causes and Consequences of Feelings.* Cambridge University Press.

Cialdini, R. B. 1998. *Influence: The Psychology of Persuasion.* Collins.

Dellaert, F.; Polzin, T.; and Waibel, A. 1996. Recognizing emotion in speech. In *Proceedings of the CMC*, 1970–1973.

Gratch, J., and Marsella, S. 2001. Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. In *Proceedings of the Fifth International Conference on Autonomous Agents.*

Gross, J. J., and Levenson, R. W. 1995. Emotion elicitation using lms. *Cognition and Emotion* 9(1):87–108.

Huber, R.; Batliner, A.; Buckow, J.; Noth, E.; Warnke, V.; and Niemann, H. 2000. Recognition of emotion in a realistic dialogue scenario. In *Proceedings of INTERSPEECH (ICSLP)*, 665–668.

Laurel, B. 1986. *Toward the Design of a Computer-Based Interactive Fantasy System.* Ph.D. Dissertation, Drama department, Ohio State University.

Lazarus, R. S.; Speisman, J. C.; Mordkoff, A. M.; and Davison, L. A. 1962. A laboratory study of psychological stress produced by motion picture film. *Psychological Monographs* 76.

Lisetti, C. L., and Nasoz, F. 2004. Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP Journal on Applied Signal Processing* 2004(11):1672–1687.

Liu, H.; Lieberman, H.; and Selker, T. 2003. A model of textual affect sensing using real-world knowledge. In *Proceedings of Intelligent User Interfaces (IUI) 2003*, 125–132.

Marsella, S. C.; Pynadath, D. V.; and Read, S. J. 2004. Psychsim: Agent-based modeling of social interactions and influence. In *Proceedings of the International Conference on Cognitive Modeling*, 243–248.

Mateas, M., and Stern, A. 2005. Structuring content in the facade interactive drama architecture. In *Proceedings of Artificial Intelligence and Interactive Digital Entertainment (AIIDE05)*.

Mateas, M. 1999. An Oz-centric review of interactive drama and believable agents. In Woodridge, M., and Veloso, M., eds., *AI Today: Recent Trends and Developments. Lecture Notes in AI 1600*. Berlin, NY: Springer. First appeared in 1997 as Technical Report CMU-CS-97-156, Computer Science Department, Carnegie Mellon University.

McHugo, G. J.; Smith, C. A.; and Lanzetta, J. T. 1982. The structure of self-reports of emotional responses to film segments. *Motivation and Emotion* 6:365–385.

Morency, L. P.; Sidner, C.; Lee, C.; and Darrell, T. 2007. Head gestures for perceptual interfaces: The role of context in improving recognition. *Artificial Intelligence* 171(8–9):568–585.

Morency, L. P.; Christoudias, C. M.; and Darrell, T. 2006. Recognizing gaze aversion gestures in embodied conversational discourse. In *International Conference on Multimodal Interactions (ICMI06)*.

Nelson, M. J., and Mateas, M. 2008. Another look at search-based drama management. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI-08)*.

Nelson, M. J.; Roberts, D. L.; Isbell, C. L.; and Mateas, M. 2006. Reinforcement learning for declarative optimization-based drama management. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-06)*.

Neviarouskaya, A.; Prendinger, H.; and Ishizuka, M. 2007. Textual affect sensing for sociable and expressive online communication. In *ACII*, 218–229.

Osherenko, A., and Andr, E. 2007. Lexical affect sensing: Are affect dictionaries necessary to analyze affect? In *Proceedings of Affective Computing and Intelligent Interaction (ACII)*. Springer.

Osherenko, A. 2008. Towards semantic affect sensing in sentences. In *Proceedings of AISB 2008 Convention on Communication, Interaction and Social Intelligence*.

Philippot, P. 1993. Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition and Emotion* 7:171–193.

Read, J. 2004. Recognising affect in text using pointwise-mutual information. Masters thesis, University of Sussex.

Roberts, D. L., and Isbell, C. L. 2008. A survey and qualitative analysis of recent advances in drama management. *International Transactions on Systems Science and Applications, Special Issue on Agent Based Systems for Human Learning* 4(2).

Roberts, D. L.; Nelson, M. J.; Isbell, C. L.; Mateas, M.; and Littman, M. L. 2006. Targetting specific distributions of trajecotries in MDPs. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI-06)*.

Roberts, D. L.; Bhat, S.; Clair, K. S.; and Isbell, C. L. 2007. Authorial idioms for target distributions in ttd-mdps. In *Proceedings of the 22st Conference on Artificial Intelligence (AAAI-07)*.

Roberts, D. L.; Isbell, C. L.; Riedl, M. O.; Bogost, I.; and Furst, M. L. 2008. On the use of computational models of influence for interactive virtual experience management. In *Proceedings of the First International Conference on Interactive Digital Storytelling (ICIDS-08)*.

Si, M.; Marsella, S. C.; and Pynadath, D. V. 2005. Thespian: An architecture for interactive pedagogical drama. In *International Conference on Artificial Intelligence in Education*.

Si, M.; Marsella, S. C.; and Pynadath, D. V. 2006. Thespian: Modeling socially normative behavior in a decision-theoretic framework. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents*.

Stiefelhagen, R. 2002. Tracking focus of attention in meetings. In *In Proceedings of the International Conference on Multimodal Interfaces*.

von Ahn, L.; Kedia, M.; and Blum, M. 2006. Verbosity: A game for collecting common-sense knowledge. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 75–78.

von Ahn, L.; Liu, R.; and Blum, M. 2006. Peekaboom: A game for locating objects in images. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 55–64.

von Ahn, L. 2006. Games with a purpose. In *IEEE Computer Magazine*, 96–98.

Weyhrauch, P. 1997. *Guiding Interactive Drama*. Ph.D. Dissertation, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA. Technical Report CMU-CS-97-109.

Yacoub, S.; Simske, S.; Lin, X.; and Burns, J. 2003. Recognition of emotions in interactive voice response systems. In *Proceedings of Eurospeech*, 1–4.