Towards an Empathizing and Adaptive Storyteller System

Byung-Chull Bae[1], Alberto Brunete[2], Usman Malik[3], Evanthia Dimara[4], Jermsak Jermsurawong[5], Nikolaos Mavridis[5]

[1]IT University of Copenhagen, Denmark, [2]Carlos III University, Spain, [3]National University of Sciences and Technology, Pakistan,
[4]Université Paris-Sud, France, [5]New York University Abu Dhabi, UAE
[1]byungchull.bae@gmail.com, [2]alberto.brunete@gmail.com, [3]usman.malik88@gmail.com, [4]evanthia.dimara@gmail.com,
[5]{jj1192, nikolaos.mavridis}@nyu.edu

Abstract

This paper describes our ongoing effort to build an empathizing and adaptive storyteller system. The system under development aims to utilize emotional expressions generated from an avatar or a humanoid robot in addition to the listener's responses which are monitored in real time, in order to deliver a story in an effective manner. We conducted a pilot study and the results were analyzed in two ways: first, through a survey questionnaire analysis based on the participant's subjective ratings; second, through automated video analysis based on the participant's emotional facial expression and eye blinking. The survey questionnaire results show that male participants have a tendency of more empathizing with a story character when a virtual storyteller is present, as compared to audio-only narration. The video analysis results show that the number of eye blinking of the participants is thought to be reciprocal to their attention.

Introduction

Storytelling is a narrative communication between a storyteller and a listener. In face-to-face storytelling, the storyteller can infer whether the listener is paying attention to the story from the listener's responses or backchannels. The backchannels include verbal responses (e.g., acknowledgement tokens such as yeah, uh huh, or mm hm) (Drummond & Hopper 1993) and nonverbal responses (e.g., head nodding, eye blinking, or smiles). When the negative backchannels (e.g., head down or blank expression from boredom throughout the storytelling) are continuously recognized, an effective storyteller will change his or her narration technique to capture the listener's attention. While the storyteller's narration techniques will be various depending on the listener profiles (e.g., age, education, preferences, etc.), emotional expression (either verbal or nonverbal) is one of the common qualities that effective storytellers have.

Several studies on emotionally expressive storytelling have been conducted using virtual agents (Silva *et al*. 2001, Charles *et al*. 2007, Bleackley *et al*. 2010, Yan & Agada 2010). Papous the virtual agent (Silva *et al.* 2001) could tell stories with human-like expressions by using behavior, scene, illumination and emotion tags in the input text. In Interactive Storytelling (IS), the story can be adapted according to the user's emotions (Charles *et al*. 2007). Blom and Beckhaus (2005) have also proposed the concept of emotional storytelling in IS, using explicitly parameterized reader's emotions for story progression. The notion of the reader's anticipated emotion tracking in their paper has similarity with our approach, but ours is focused rather on the listener's attention recognition and storyteller's empathizing telling by using the emotions of the story characters.

Storyteller agents or systems can detect the listener's nonverbal responses using various sensor devices. Our interest in this paper is to detect the listener's nonverbal backchannels as a sign of attention or engagement in the story. The detection of positive backchannels is an indication of the listener's satisfaction or engagement in the story. In this case the storyteller system will continue to tell the story with the current storytelling mode or technique. The detection of negative backchannels over the specified threshold, on the other hand, could be a sign of the listener's disliking or inattentiveness, implying the change of the current storytelling mode. We hypothesize that the listener's valence to the storytelling (either positive or negative) could be detected through the backchannels.

System Design and Experiment

Overall System Architecture

Our system architecture under development is illustrated in Figure 1, consisting of four main components - Discourse Manager, Narrative Discourse Generator, Attention Cue Generator, and Attention Detector. The system takes a text story file with emotion tagging as input and generates an emotionally expressive oral narrative discourse. Based on the analysis of the listener's attentiveness in real-time, the system will also generate nonverbal attention cues and will modify narrative discourse techniques. To generate narrative discourse and test the necessary storytelling techniques, we use Greta, an embodied conversational agent (Poggi *et al.* 2005), as a virtual storyteller. We employ two software toolkits for the recognition of the listener's attentiveness in real-time – FaceAPI [1], a commercial face-tracking software by Seeing Machines, and SHORE^{TM}(Sophisticated High-speed Object Recognition Engine) [2], a live facial emotional expression analyzer by the Fraunhofer Institute for Integrated Circuits in Erlangen, Germany. Although we currently use a virtual avatar as an oral storyteller for our system development, we plan to gradually import the emotionally expressive storytelling techniques to a physical robot.

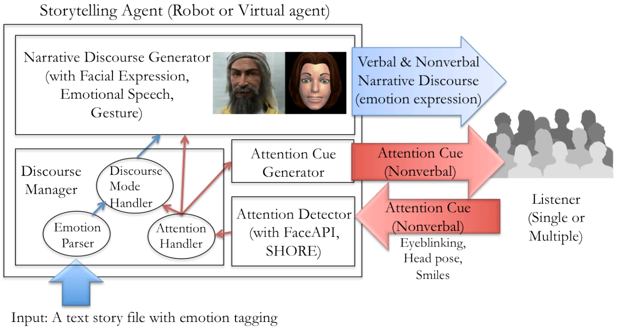
Storytelling Agent (Robot or Virtual agent)
Narrative Discourse Generator (with Facial Expression, Emotional Speech, Gesture)
Verbal & Nonverbal Narrative Discourse (emotion expression)
Discourse Manager
Discourse Mode Handler
Emotion Parser
Attention Handler
Attention Cue Generator
Attention Cue (Nonverbal)
Attention Detector (with FaceAPI, SHORE)
Attention Cue (Nonverbal)
Eyeblinking, Head pose, Smiles
Listener (Single or Multiple)
Input: A text story file with emotion tagging

Figure 1. Overall System Architecture

Experiment

We conducted a pilot study to examine the overall validity of our system design. We hypothesized that a virtual storyteller's presence would facilitate the listener's story liking, engagement, and empathizing with a story character. We also hypothesized that the listener's attentiveness could be detected through the analysis of the listener's facial emotion expression and eye blinking.

Material and Procedure

As the story material we chose a short story titled The Cracked Pot (two minutes long including about 200 words). The story is about a cracked pot that considers itself useless because of its inherent crack, a critical flaw for the pot holding water, but finally realizes its own beauty and usefulness as it is. While the story is short, it is emotional (including emotions such as shame, proud, empathy, happiness, sadness, etc.) and enjoyable regardless of age. In the story a heterodiegetic narrator (i.e., the narrator who is not present in the storyworld as a character) narrates a story of three characters (a woman, a perfect pot, a cracked pot) with omniscient point of view.

To obtain an objective emotion tagging, five of the authors of this paper (who coincidentally have cross-cultural backgrounds – Korea, Spain, Greece, Pakistan, and Thailand) carefully read the story several times and individually tagged the possible character emotions sentence by sentence. The emotion category was limited to Ekman's six basic emotions (Happiness, Sadness, Anger, Fear, Disgust, and Surprise) with intensity range from 0 (Not all all) to 10 (Extreme). The tagged data were collected and averaged with the confidence ratio based on the number of the responses.

A total of 20 subjects (10 women, 10 men), who were students, staffs, and researchers in New York University Abu Dhabi, were volunteer participants in the experimental study. Their ages ranged from 18 to 60 years old. Each participant was arbitrarily assigned, considering the gender ratio, to one of the two subject groups. Most of the subjects used English as a foreign language but did not have difficulty in understanding the story in English. The participants in one group (Group A) individually listened to the pre-recorded audio story which is narrated by a human storyteller without any video stimuli; the participants in the other group (Group B) also listened to the same audio story individually but with video where Greta expresses her emotion using facial expression according to the pre-recorded audio. We decided to record the story by a human female voice because story narration using the built-in TTS (Text-To-Speech) in Greta was not comprehensible enough for non-English speakers in terms of voice tone and speed. The facial expressions of the participants in Group B were recorded during the storytelling and analyzed later offline. After the story was over, the participants were asked to provide ratings on a 7-point scale ranging from not at all (1) to very much (7) about their story appreciation. The survey questionnaire included six closed questions regarding the story in terms of story liking, engagement, and empathizing with the main story character (i.e., the cracked pot) and several open questions regarding the emotions of story characters and the narrator.

[1] http://www.seeingmachines.com/product/faceapi/
[2] http://www.iis.fraunhofer.de/en/bf/bsy/produkte/shore/

Results

The mean ratings of the three story appreciation factors (liking, engagement, and empathizing) in two groups (Audio storytelling vs. Video storytelling using Greta) are not considerably different. Instead, some gender differences are noticed. The difference of mean ratings between male and female participants is trivial in the subject group B (Video storytelling using Greta) – less than 3%. However, the difference in the subject group A (Audio storytelling) is noticeable – more than 20%. Particularly the mean ratings of the two questions related to empathizing (e.g., How sorry did you feel for the bad pot in the beginning of the story?; How happy did you feel for the bad pot at the end of the story?) between male and female participants show the largest difference (27.1%) in the audio only storytelling.

The emotional analysis made with SHORE shows that different emotions are measured from the subject's faces while viewing the video. Overall the emotions expressed by Greta (as the narrator of the story) and the emotions recognized from the listeners share some common similarities and differences at the same time. In both happiness is gradually increasing along the story, and there is little surprise. On the other hand, however, there is a lot of sadness in Greta but not in the subjects, and a lot of anger in the subjects but not in Greta. Both are decreasing as the story progresses. This can be explained as follows: if we consider that anger and sadness are related in some ways, it is likely that the software mixed up both emotions. However, it can be derived that Greta's emotions have a correspondence in the subject's emotions.

Regarding the subject's face detection analysis, the average number of blinking of each participant while watching the video with Greta's storytelling which is thought to be reciprocal to their attention. We divided the story into three parts and compared the average number of blinkings of participants in each of them. As the video storytelling progresses the number of blinkings decreases, corresponding to an increase in subject's attention as they get involved in the story. The average number of blinkings undergoes a local minimum at the turning point (i.e., climax) of the story. At the end of the story the attention decreases, and the blinking ratio increases again towards its finale.

Conclusion and Discussion

In this paper we propose the architecture for an empathizing and adaptive storyteller system under implementation. The results of our pilot study using Greta, an embodied conversational agent, showed that the use of emotional virtual storyteller could enhance the listener's (especially male listener's) experience of empathizing with a story character. The study results also showed that the recognition of the listener's emotional facial expression and eye blinking could be effective for detecting the listener's attention or empathizing with a story character.

While Greta is an efficient tool for developing an emotional virtual storyteller, its text-to-speech system seems to have a room for enhancement. Especially for listeners who speak English as a foreign language, clear and articulate pronunciation as well as appropriate tone for storytelling is required.

As future work we consider using an emotionally expressive humanoid robot as storyteller. Unlike the virtual agent storyteller, the robot storyteller will elicit a strong sense of presence. We expect that the robot storyteller's sense of presence will be persuasive in adaptive storytelling where the robot storyteller and the listener can communicate with each other through backchannels. We also plan to conduct a similar experimental study with a different story narrated by a male virtual storyteller.

Acknowledgements

We would like to thank the developers of SHORE^{TM} and Greta for their generous permission to let us use their respective software toolkits.

References

Bleackley, P., Hyniewskab, S., Niewiadomski, R., Pelachaud, C. and Price, M. 2010. Emotional Interactive Storyteller System. In Proc. International Conference on Kansei Engineering and Emotion Research, Paris, France.

Blom, K.J. and Beckhaus, S. 2005. Emotional Storytelling. In Virtual Reality Conf., Workshop "Virtuality Structure" 23-27

Charles, F., Lemercier, S., Vogt, T., Bee, N., Mancini, M., Urbain, J., Price, M., Andre, E., Pelachaud, C. and Cavazza, M. 2007. Affective Interactive Narrative in the CALLAS Project. In Proceedings of ICVS, Saint-Malo.

Drummond, K. and Hopper, R. 1993. Back Channels Revisited: Acknowledgement Tokens and Speakership Incipiency. Research on Language and Social Interaction, 26 (2).

Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., and De Carolis, B. 2005. Greta: A Believable Embodied Conversational Agent. Multimodal Intelligent Information Presentation. Sotck, O. and Zancanaro, M. (Eds).

Silva, A., Vala, M., and Paiva, A. 2001. Papous: The Virtual Storyteller. In Proceedings of Intelligent Virtual Agents.

Yan, J. and Agada, R. 2010. Life-like Animated Virtual Pedagogical Agent Enhanced Learnin. 2010. J. of Next Generation Information and Technology, 1(2), 4-12.