Facial Expressions of Appraisals displayed by a Virtual Storyteller for Children

Nesrine Fourati[1], Adeline Richard[2], Sylvain Caillou[1], Nicolas Sabouret[1], Jean-Claude Martin[1], Emilie Chanoni[2], Celine Clavel[1]

[1] LIMSI, CNRS (UPR3251), Université Paris-Saclay, Orsay, F-91405, France
{fourati, sylvain.caillou, nicolas.sabouret, jean-claude.martin, celine.clavel}@limsi.fr
[2] PSY-NCA, Université de Rouen, 76000 Rouen, France
{adeline.richard, emilie.chanoni}@etu.univ-rouen.fr

Abstract. In this paper, we present a framework for an expressive virtual storyteller for children. Our virtual storyteller displays facial expressions of appraisals related to story events. The facial expressions are animated jointly with deictic gestures towards graphical elements representing story events. We describe a preliminary study that we conducted with 23 children. We discuss the impact of facial expressions of appraisals on children's memorization of story events, their perception of characters' appraisals, their subjective perception of the virtual storyteller and more generally how emotion combines with joint attention.

Keywords: Expressive virtual storyteller, Facial expression, Appraisal

1 Introduction

In addition to its entertaining values, storytelling supports children's development of cognitive, communicative, linguistic and literacy skills as well as logical thinking [18,5]. Nonverbal behaviors displayed by a human storyteller, including facial expressions of emotions, intonation, and communicative gestures, do provide important information about the meaning of the story as well as the personal interpretation of this story by the storyteller. Storytelling with virtual agents and robots has emerged in the nineties. It allows children to experience new forms of interaction and possibly to increase their enjoyment [12]. Several studies highlight the importance of interactive storytelling frameworks to better engage children with the story narrated by a virtual agent and to foster their creativity during storytelling [24,12]. Joint Attention (JA) behavior also allows interacting with the partner in order to share a particular experience through coordinated attention to an object (or event) with mutual affect [13]. JA behavior can be achieved by the means of verbal (e.g. "look at this object") and non-verbal behaviors (e.g. gaze direction, pointing gestures) [21]. Besides, contextual expressive capabilities of a virtual storyteller are important for children [12]. Several studies propose virtual agents endowed with socio-emotional skills [20] [11]. Studies observe that endowing virtual agents with the ability to express emotions and social attitudes might improve the quality of the interaction

with the user compared to non expressive agents [14] [19] [2]. Courgeon et al. [9] adapted a computational model of emotion based on an appraisal theory in the case of a reversi game (but not in a storytelling task). Whereas the expression of emotion and joint attention initiation seem to be major requirements for a virtual storyteller for children, few virtual storytellers do exploit these two communicative functions together as a function of the current context [4]. Besides, the joint effect of expressions of emotions and joint attention initiation on children's experience and understanding of a story remains understudied. In this paper, we explain how we combine emotion expression and joint attention initiation within a single framework of a virtual storyteller for children. Our first aim is to study the contribution of emotion expression when combined with deictics. The present paper focuses on the modeling of emotion expression during a storytelling task and its impact on children's memorization of the story's emotional events, their evaluation of the characters' emotional states and their subjective perception of the storyteller. In the next section, we present related work on artificial storytellers. We then present in section 3 our methodology for animating such an artificial storyteller capable of expressing emotions related to a story and displaying deictic behaviors. In section 4 we present a preliminary study with 23 children. Section 5 discusses future work including future evaluations and how such work can be used to inform the design of autonomous expressive storytellers.

2 Related Work

Storytelling for children recently received a growing interest in the Human-Computer Interaction community. Several interactive storytelling frameworks, applications, and devices are proposed such as interactive toys [7], interactive books, robots [15] and computer games characters [3]. Studies suggest positive impacts of exploiting virtual agents compared to a text displayed on a screen [26]. Several studies focus on the design of interactive storytelling frameworks to establish a natural interaction [7] [3]. However, expressing and acting emotions sounds important to offer a positive storytelling experience [25]. Besides, the expression of emotion can have a considerable impact on the believability of an artificial storyteller [8]. Recently, there has been a growing interest on the development of storytellers endowed with capabilities to express emotions. The need to generate expressive and emotional speech was considered for the design of an expressive character-based storyteller in [8] and [25]. In [23], the authors present their expressive virtual storyteller called Papous. A text-to-speech framework is used and expressive parameters are employed to endow the virtual storyteller with the ability to express emotion through prosody. The authors claim that the agent's voice still seems synthetic despite efforts to make it sound expressive [23]. Other studies focus on bodily expressions of emotions and moods during storytelling for example using the NAO robot [15]. Only a few studies consider the combined impact of emotion expression and joint attention initiation on the experience of Storytelling with children [4].

3 Methodology

Our expressive virtual storyteller is based on the MARC virtual agent framework (Multimodal Affective and Reactive Characters) [10]. The female agent, named Mary, is used for our study (see Figure 1). We selected a French story called "Le ballon perché" because of its pedagogical relevance and for its rich expressive content. The story is about three children playing football just before the class starts. The ball gets stuck on a roof. The children try to get the ball down by throwing their personal items on the roof. However, the situation gets even worse since all of the items also remain on the roof. The children have to go back to the classroom and they face several difficult situations involving their missing items, for which they find workarounds. Finally, a storm blows all the items and the ball off the roof. This story is full of unexpected situations for young children, while happening in a familiar and realistic frame.

Fig. 1. Our expressive virtual storyteller displaying facial expressions of appraisals and performing a pointing gesture.

Since our goal is to endow our virtual storyteller with the ability to drive an interactive and expressive narration, the original story content has been revised by researchers in psychology to include interactive utterances in addition to narrative utterances (160 utterances totally). Thus, we distinguish two categories of utterances: 1) narrative utterances (56.25% of all the utterances), and 2) interactive utterances (43.75% of all the utterances). Narrative utterances correspond to the original story content. Interactive utterances are useful to maintain engagement during the narration task. We define different forms of interactive utterances; Interrogative, Affirmative and Joint Attention (JA) initiation utterances. An Interrogative utterance refers to a question which aims to maintain the child's engagement (e.g. "do you think he is going to make it?"). An Affirmative utterance refers to the introduction and the conclusion phases of the story (e.g. "Hello my name is Mary") or to the transition between the child's answer and a narrative utterance (e.g. "That's what Im thinking too"). JA utterances are used to drive the child's attention to events or objects related to the story (e.g. "look at this teacher").

During narration task, JA utterances are coupled with nonverbal behaviors to direct the child's attention to a character/ event/ object in the story images. For each JA utterance, the agent's gaze and his head movement are oriented toward

the image. In addition to the gaze and head movement, we also include either torso movement or pointing gesture. The former is used to stress the presence of a new event. The latter is employed to point a particular object or character. To maintain the child's attention and to be able to initiate joint attention, we imported graphics from the story book into the 3D virtual environment and displayed besides the virtual storyteller (see Figure 1).

In the present study, we aim to endow our agent with the ability to display facial expressions (see section 3.1). As audio recording results in a more natural storyteller compared to a synthetic voice, we asked a (female) psychology student to record the speech of each utterance in a spontaneous and an expressive way. 160 audio files were collected (one per utterance). We synchronized lips motion provided from the utterance content based on the JSAPI text to speech tool with the audio recording of narration.

3.1 Facial expressions

Different theories were proposed to model the morphological and dynamic characteristics of facial expressions of emotions. They are commonly grouped in three approaches: Categorical, Dimensional and Appraisal approaches. Appraisal theory is one of the most influential theory within affective computing [17]. It has been widely used as the basis for several computational models of emotion [16]. In our work, we adopt the Component Process Model (CPM), an appraisal theory, for its dynamic ability to evaluate and describe an adaptive reaction to a story event [22]. The five appraisal variables (also called "Stimulus Evaluation Checks") described in [22] have been selected and considered for our study: Novelty, Intrinsic Pleasantness, Goal/Need Conduciveness, Coping Control and Coping Power. Norm/Self compatibility check has not been considered in our study for the lack of insights regarding the activation of the corresponding Action Units. A mapping between appraisal variables and Action Units from the FACS (Facial Action Coding System) was proposed in [22]. A computational model of these five appraisal variables has been developed within the MARC platform [9]. The mapping between each appraisal variable and the associated facial action units is mainly based on the work described in [22].

A manual annotation step was performed to attribute the appropriate appraisal evaluation for each utterance of the story. As such, the agent will display the appropriate facial expression according to each utterance based on appraisal annotation. Five researchers (the authors) from Affective Computing and Psychology were asked to perform this annotation step. For each utterance, each annotator was asked to explain the viewpoint used to evaluate the appraisal variables (e.g. Narrator or Story Character viewpoint) and to attribute the appropriate discrete value of each appraisal variables based on a 5-points scale (-1, -0.5, 0, 0.5, 1). The annotators were provided with the definition of each appraisal variable. As we aim to obtain a single annotation of appraisal to animate our virtual agent during the storytelling task, the annotators met to discuss the differences in their annotation in order to reach an agreement on the appropriate

annotation of appraisal variables for each utterance. At the end of the annotation process, 57.50% of the utterances were evaluated as expressive (i.e. at least one appraisal variable was not set to 0). The other utterances were defined as non-expressive (i.e. no facial expressions should be displayed). However, we assume that endowing a virtual agent with smiles during the storytelling task can create a friendly atmosphere, which is of high importance during an interaction with children. For this reason, during non-expressive utterances, we endowed our virtual agent with a smiling face based on the morphological characteristics of an *amused* smile [20] (i.e. 6 (Cheek raiser), 12 (Lip corner puller) and AU25 (lips parted)). We chose to display facial expressions (facial expressions of appraisal and amused smile respectively for expressive and non-expressive utterances) during the whole spoken utterance based on the results described in [6]. These results indicate that facial expressions seem to be perceived as more intense and more realistic when displayed during the whole spoken utterance [6].
In the next section, we present a preliminary study that provides first insights into the impact of facial expression of appraisal during storytelling task on the cognitive skills of children and on their ability to evaluate emotional states based on an appraisal model. In the present study, we do not investigate the impact of Joint Attention initiation on these components. This research question will be addressed in a future work.

4 Exploratory study

We conducted an exploratory study to explore how facial expressions of appraisals may affect children cognitive processes and evaluation of characters' appraisals. The narration of the story lasts around 10 minutes. Twenty three children contacted via a recreation center participated in our experiment (10 female). One recording session was performed per child. During the whole recording session, the child was accompanied with an adult. The children were between 6 and 10 years old. Twelve children were asked to listen to the story narrated by our virtual agent in expressive condition, while eleven children were asked to listen to the story in non-expressive condition. In the expressive condition, our virtual agent displayed facial expressions of appraisals for each expressive utterance based on the annotation described in section 3.1. In the non-expressive condition, our virtual agent displayed an amused smile all along the storytelling task (see section 3.1). The child was free to make spoken comments during the whole storytelling task. At the end of the story, the child was asked to fill a survey with the help of an adult. The videos of the children were also recorded for later study of their interaction with the virtual storyteller. The parents of each child gave informed consent and agreed that their children' answers to the survey as well as their videos could be used and published for research purposes.

4.1 Measures

Our survey is composed of three parts.

- Twenty-three questions used to assess the memorization of story's events. The answers to these questions were converted into numerical values belonging to the interval [0,1] (0 means the answer is wrong, 1 means the answer is right, between 0 and 1 means the answer is partially right).

- Eight questions used to examine how children evaluate several appraisal variables. They were inspired from the Geneva Appraisal Questionnaire [22] and adapted to children's cognitive skills . Each question is associated to a specific story event illustrated by the corresponding story's image. For instance, the second question is associated to the story image of one character trying to throw a personal item onto the roof in order to get the ball down. For each question, the child was asked to evaluate 5 appraisal variables along a 5-point scale from "totally disagree" to "totally agree".. The appraisal variables are Novelty, Intrinsic pleasantness, Goal Conduciveness, Coping Control and External Causation (i.e. Compatibility with social norms) [22]. Five assertions were used to help the child evaluate each appraisal variable. For instance, the assertion corresponding to Novelty check was "It's something that does always happen."

- Fifteen questions used to explore how the virtual storyteller is perceived (e.g. nice, friendly, pleasant...). The content of these subjective questions was inspired by the French Translation of Bartneck's Godspeed Questionnaire[1]. The answer was given by the means of a 5-points scale.

4.2 Preliminary results

Based on the measures described above, we discuss the impact of facial expressions of appraisals on 1) the children's memorization of the story's events, 2) their evaluation of appraisals and 3) their appreciation of the virtual storyteller.

Do facial expressions of appraisals impact children's memorization of the story? In order to study the impact of facial expressions of appraisals, the answers of the 21 questions that address the memorization of story's events were averaged to obtain a measure of the overall memorization of story's event for each subject. A t-test was applied to study the effect of facial expressions on the overall memorization of story' events. No significant effect was found (see Table 1). However, we observed that children tended to slightly better memorize story's events in the expressive condition (see Table 1). Pointing gestures were used to initiate the joint attention for the events addressed in eleven questions of memorization (out of 21). Indeed, the pointing gestures used to initiate the joint attention for these events could affect the child's memorization regardless of the facial expressions.

The age of our participants varies between 6 and 10 years old (std=0.94, mean=8, 14 children under 8 years old). A t-test was conducted to explore the effect of the age on the overall memorization of events. A significant difference was found between the groups of children under and above 8 years old ($p<0.001$). We observed that older children tended to better memorize story's events than younger children (see Table 1). This result might explain why no significant difference was found between the memorization of events in expressive and non-expressive conditions. Given the strong effect of age on the memorization of

Table 1. The effect of facial expressions and of the child's age on the overall memorization of story events.

	t-test	Mean	Std	Min	Max
Non-Expressive Storyteller	p=0.58	0.65	0.15	0.44	0.88
Expressive Storyteller		0.68	0.14	0.46	0.93
Younger (under 8 years old)	p<0.001	0.60	0.12	0.44	0.81
Older (above 8 years old)		0.78	0.09	0.68	0.93

the story's events, one may conclude that more participants on a more focused age range will be needed to investigate the effects of emotion expressions on children's cognitive skills.

For the sake of simplicity, we will respectively call older and younger children the subjects who are above and below 8 years old (the mean age of our participants). We found that young children tended to better memorize events related to the story content and reported in joint attention initiation utterances in non-expressive condition ($p<0.05$). Indeed, it seems that facial expressions disturb young children's memorization of story's events when the agent initiate joint attention for events related to the story content. However, the same was not observed for the old children group who tended to memorize story's events regardless of the presence or absence of facial expressions of appraisals. This difference might can be explained by the content of the story and the interactive scenario that we created which might not be relevant for children above 8 years old. This difference in the results suggests to conduct an evaluation only with young children (e.g. less than 8 years old).

Do facial expressions of appraisals impact children's evaluation of appraisals? We applied a t-test for each of the 40 answers of appraisals evaluation to study the impact of facial expressions on children's evaluation of appraisals (8 questions * 5 appraisal variables, see section 4.1). No significant effect was found, which could also be related to the variety of age. We explored the effect of facial expressions on their evaluation of appraisals separately for younger and older group of children. A serial of t-test was conducted for this purpose for the 40 answers. We found that facial expressions of appraisals displayed during narration impacted young children's evaluation of Novelty across different events($p<0.05$). Overall, younger children tended to evaluate different events as more Novel when the virtual agent displayed facial expressions of appraisals. As such, the presence of facial expressions of appraisals accentuates the evaluation of Novelty. For instance, the event related to "Standing in line before class starts" was evaluated as expected (e.g. always happening) in the non-expressive condition. It was evaluated as significantly less expected in the expressive condition. In the expressive condition, the virtual storyteller displayed facial expressions associated to non-expected, unpleasant and obstructive to the children's goal since the children did not get the ball down in time and their items were still stuck on the roof. Another similar significant finding was observed for the evaluation of the "Ex-

ternal causation" variable with younger children. In the expressive condition, younger children tended to evaluate a few events as being not in line with social norms. In the non-expressive condition, they tended to evaluate these events as in line with social norms. For example, the event "Standing in line before class starts" tended to be evaluated as less respecting social norms in the expressive condition where the virtual storyteller expressed the emotional states of the characters being in an awkward situation (the characters had to go to the classroom without their personal items). Besides, we compared the children's evaluation of appraisal variables with the annotators' manual annotation. We found that the children's evaluation of these appraisal variables can vary according to the viewpoint that one takes, which has been already be reported by the annotators (see section 3.1). For example, the event "the teacher is screaming" would be surprising in a general situation, but might be expected in the case of our story when the children were standing much too close to the window. In the expressive condition, we found that the children's evaluation of these appraisal variables (for the corresponding particular events) tended to be correlated with the annotators' evaluation according to the *story content*. Besides, children's evaluation of these appraisal variables in the non-expressive condition tended to be correlated with our annotation in absolute terms. As such, it seems that facial expression of appraisals helped children to refer to the story content to evaluate appraisal variables. When no facial expression was displayed, children tended to evaluate events according to a non story-specific point of view. Overall, these findings suggest potential impacts of displaying facial expressions of appraisals on the evaluation of appraisal variables by children.

Do facial expressions of appraisals impact children' appreciation of the virtual storyteller? A serial of t-test was applied on each of the 15 questions that assess the children's subjective evaluation of the agent. No significant difference was found between their subjective evaluation in expressive and non-expressive condition. Similarly to the results reported above for the impact of facial expressions of appraisals on children's memorization, we observed that children's appreciation of the virtual storyteller tend to depend on their age. Young children tended to evaluate the virtual storyteller as being more or less human (means=2.6 with with 0 meaning "machine" and 4 "human") while old children tended to evaluate the agent as being less human (mean=1.7). However, old children tended to consider the agent as being more likeable (means=3.6) than young children (means=3.07). This result might be explained by the fact that young children tended to perceive the agent as being more alive and thus had high expectations about it. Old children might have been more aware that the agent was a computer software and thus were more tolerant to graphical flaws.

5 Conclusion and future directions

Our framework supports the expression of appraisals of the story content and the initiation of joint attention during the storytelling task. The Component

Process Model was used to drive the virtual agent's Action Units to display corresponding facial expressions of appraisals. An exploratory study with 23 children was conducted to explore the impact of these facial expressions of appraisal on children' perception and understanding of the story. Understanding and appreciation of the virtual storyteller depended on the age of the child. Young children's cognitive skills and evaluation of the appraisals were impacted by the facial expressions of appraisals displayed by the virtual storyteller. In the short term, we will conduct an experimental study with a larger pool of participants with a more homogeneous age range. We will also use an eye-tracker to record the direction of the children's gaze. This will enable us to explore the frequency and the duration of children's gaze toward the agent's face vs. towards the story graphics. In the long-term, we aim to add bodily expressions of emotion. Finally, the evaluation of the appraisal variables in the present work has been performed based on a manual annotation and a human evaluation of the story events and of the agent viewpoint. An automatic evaluation of appraisal variables would represent an important step towards fully automatic virtual storytellers.

Acknowledgments. The authors would like to thank Matthieu Courgeon for his help in the development of our virtual storyteller. A special thanks goes to the children, their parents and the director of the recreation center. Part of this work was funded by the Agence Nationale de la Recherche (ANR), project NARECA (ANR-13-CORD-0015).

References

1. Bartneck, C., Croft, E., Kulic, D.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics 1(1), 71–81 (2009)
2. Beale, R., Creed, C.: Affective interaction: How emotional agents affect users. International Journal of Human-Computer Studies 67(9), 755—776 (2009)
3. Bernardini, S., Porayska-Pomsta, K., Smith, T.J.: ECHOES: An intelligent serious game for fostering social communication in children with autism. Information Sciences 264, 41–60 (2014)
4. Bosse, T., Pontier, M., Siddiqui, G.F., Treur, J.: Incorporating emotion regulation into virtual stories. In: Intelligent Virtual Agents, 7th International Conference, IVA 2007, Paris, France, September 17-19, 2007, Proceedings. pp. 339–347 (2007)
5. Bruner, J.: The Narrative Construction of Reality. Critical Inquiry 18(1), 1–21 (1991)
6. Buisine, S., Wang, Y., Grynszpan, O.: Empirical investigation of the temporal relations between speech and facial expressions of emotion. Journal on Multimodal User Interfaces 3(4), 263–270 (2009)
7. Cassell, J., Bers, M.U.: Interactive storytelling systems for children: using technology to explore language and identity. Journal of Interactive Learning Research 9, 183–215 (1999)
8. Cavazza, M., Pizzi, D., Charles, F.: Emotional input for character-based interactive storytelling. Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems pp. 313–320 (2009)

9. Courgeon, M., Clavel, C., Martin, J.C.: Modeling Facial Signs of Appraisal During Interaction: Impact on Users' Perception and Behavior. AAMAS 59(1), 765—772 (2014)
10. Courgeon, M., Martin, J.C., Jacquemin, C.: Marc: a multimodal affective and reactive character. In: Proceedings of the 1st Workshop on AFFective Interaction in Natural Environments (2008)
11. De Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., De Carolis, B.: From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. Int. journal of human-computer studies 59(1), 81—118 (2003)
12. Fridin, M.: Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. Computers & Education 70, 53–64 (2014)
13. Kiser, L.J., Baumgardner, B., Dorado, J.: Who are we, but for the stories we tell: Family stories and healing. Psychological trauma: theory, research, practice and policy 2(3), 243–249 (2010)
14. Kramer, N.: Social effects of virtual assistants. a review of empirical results with regard to communication. Intelligent virtual agents pp. 507—508 (2008)
15. Le, Q.A., Hanoune, S., Pelachaud, C.: Design and implementation of an expressive gesture model for a humanoid robot. 2011 11th IEEE-RAS International Conference on Humanoid Robots pp. 134–140 (2011)
16. Marsella, S.C., Gratch, J.: EMA: A process model of appraisal dynamics. Cognitive Systems Research 10(1), 70–90 (2009)
17. Marsella, S.C., Gratch, J., Petta, P.: Computational Models of Emotion. In: Scherer, K.R., Banziger, T., Roesch (eds.) A blueprint for affective computing: A sourcebook and manual, pp. 21–46. Oxford University Press, Oxford (2010)
18. Miller, S., Pennycuff, L.: The Power of Story : Using Storytelling to Improve Literacy Learning. Journal of Cross-Disciplinary Perspectives in Education 1(1), 36–43 (2008)
19. Ochs, M.: Modélisation, formalisation et mise en oeuvre d'un agent rationnel dialoguant émotionnel empathique. Ph.D. thesis, Doctoral dissertation, Paris 8 (2007)
20. Ochs, M., Niewiadomski, R., Pelachaud, C.: How a virtual agent should smile? Morphological and dynamic chacteristics of virtual agent's smiles. In: Intelligent Virtual Agents. pp. 427–440 (2010)
21. Scaife, M., Bruner, J.: The capacity for joint visual attention in the infant. Nature 253(5489), 265–266 (1975)
22. Scherer, K.R.: Appraisal Considered as a Process of Multilevel Sequential Checking. In: Appraisal processes in emotion: Theory, Methods, Research, pp. 92–120. Oxford University Press (2001)
23. Silva, A., Vala, M., Paiva, A.: Papous: The Virtual Storyteller. In: de Antonio, A., Aylett, R., Ballin, D. (eds.) Proceedings of Intelligent Virtual Agents, Third International Workshop, IVA 2001. pp. 171–180. Springer (2001)
24. Theune, M., Linssen, J., Alofs, T.: Acting , Playing , or Talking about the Story : An Annotation Scheme for Communication during Interactive Digital Storytelling. 6th International Conference on Interactive Digital Storytelling pp. 132–143 (2013)
25. Theune, M., Meijs, K., Heylen, D., Ordelman, R.: Generating expressive speech for storytelling applications. IEEE Transactions on Audio, Speech and Language Processing 14(4), 1137–1144 (2006)
26. Walker, J., Sproull, L., Subramani, R.: Using a human face in an interface. Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence pp. 85—91 (1994)