How an expressive virtual storyteller for children should express emotions?

Nesrine Fourati  
LIMSI  
rue John von Neumann  
91405 Orsay  
fourati@limsi.fr

Nicolas Sabouret  
LIMSI  
rue John von Neumann  
91405 Orsay  
sabouret@limsi.fr

Adeline Richard  
PSY-NCA  
Rue Lavoisier  
76000 Rouen  
adeline.richard@etu.univ-rouen.fr

Emilie Chanoni  
PSY-NCA  
Rue Lavoisier  
76000 Rouen  
emilie.chanoni@univ-rouen.fr

Celine Clavel  
LIMSI  
rue John von Neumann  
91405 Orsay  
clavel@limsi.fr

Jean-Claude Martin  
LIMSI  
rue John von Neumann  
91405 Orsay  
martin@limsi.fr

ABSTRACT

This paper focuses on the development of an expressive virtual storyteller for children and raises the challenges encountered to endow a virtual agent with storytelling capabilities. We focus on facial expressions of emotions and we present our computational model based on an appraisal theory of emotion. We also discuss different viewpoints a storyteller can adopt to evaluate the expressive content of a story event. Besides, the paper introduces our interactive framework which aims at maintaining the attention of the child during the story.

Keywords

Appraisal theory; Emotion Expression; Expressive storyteller character

1. INTRODUCTION

Storytelling supports children’s development of cognitive, communicative, linguistic and literacy skills as well as logical thinking [25] [16] [15]. Storytelling is used as a means to help children practice their imagination, foster their creativity and their ability of interpretation, analysis and synthesis. Storytelling and the related interactions can also support relationships with adults and peers [6] [7] [26]. Besides, Storytelling for children proposes frames of reference which give information on emotions, what might cause them, depending on the current situation described by the story and on how these emotions are expressed by the storyteller. These frames of reference also allow the child to integrate the cultural norms of the expression of emotions. The activity of reading and understanding a story in an illustrated book for children is cognitively complex, requiring the ordering of thought in order to build up a mental representation which integrates both textual information, graphics and the reader’s knowledge [17]. A storyteller plays an important role to facilitate the development of a situation model. In addition to a paper book, the spoken utterances produced by a storyteller brings additional information about the meaning of the story. Nonverbal behaviors produced by a storyteller, including emotional expressions, intonation, and gestures, do provide important information about the meaning of the story as well as the personal interpretation of the story by the storyteller. Through the storyteller, the story becomes interactive. Nonverbal behaviors of children who listen to the story provide in return indicators of their levels of interest and of understanding. The storyteller can therefore regulate in real time his communication and interpretation of the story. This is a situation of co-production which takes into account the expectations, interest and level of understanding of both the children and the storyteller.

Since the nineties, as advanced technologies surround more and more our daily life, a new trend has emerged involving virtual characters in storytelling task. Storytelling with embodied virtual peers allows children to experience new forms of interaction and to increase their enjoyment [16]. Several studies highlighted the importance of interactive storytelling frameworks to better engage children with the story narrated by a virtual agent and to foster their creativity in the storytelling experiences [33, 16]. Among other features, joint attention has been identified as a fundamental characteristic of interaction. It can share focus of our partner [30]. For this reason, the design of an interactive framework generally requires deictic behaviors involving several modalities such as pointing gestures, posture shift and gaze for joint attention initialization.

In addition to such an interactive setting, an expressive narration is equally important for the engagement and for the children’s enjoyment [16]. Several studies propose virtual agents with socio-emotional skills [28] [14] [29] and observe that they might improve the quality of the interaction with the user compared to non expressive agents [20] [18] [27] [2]. For example, Courgeon et al. [13] explain how they adapted and implemented a computational model of emotion based on appraisal theories within a game task. However, computational model of emotions and the related facial expressions have been less addressed in storytelling tasks.

In our work, we focus on the modeling of emotional expression during narration with a virtual agent. Our aim is to deliver for children a storytelling experience with a virtual peer that is as “natural” as provided by a human storyteller. While it feels for us natural to comment the story events while it feels for us natural to comment the story events, we observe that they might improve the quality of the interaction with the user compared to non expressive agents [20] [18] [27] [2]. For example, Courgeon et al. [13] explain how they adapted and implemented a computational model of emotion based on appraisal theories within a game task.
non-verbal cues (such as the prosody, facial expressions and gestures), several challenges can be identified when it comes to develop an expressive virtual storyteller for children.

Whereas expression of emotion and joint attention initiation seem to be two major requirements for designing a virtual storyteller, only few systems were designed to cope with both of them [4] [1]. In this paper, we discuss the main challenges we encounter to build an expressive virtual storyteller for children and the solutions we propose. We also introduce our interactive storytelling framework based on joint attention initiation. We aim to explore how its combination with facial expressions of emotions might be perceived by children.

In the next section, we discuss existing work on storytelling with interactive and expressive artificial storytellers. We then present our methodology in section 3 for building such an artificial storyteller capable of expressing emotions and show a joint-attention behavior. It is based on annotation of the script of a story for the joint generation of expressive narration and joint attention behaviors. We explain in section 4 how the affective behavior is controlled. Section 5 concludes and discuss our future evaluation of our framework with 6-8 years-old children.

2. RELATED WORK

Storytelling for children has recently received a growing interest in the Human Computer Interaction community. Several interactive storytelling frameworks, applications, and devices have been introduced in previous works such as interactive toys [10], interactive books, robots [19] and computer games characters [3]. Previous studies showed the importance of employing talking faces and humanoid virtual agent during an interaction compared to text display on a screen [36].

In [10], the authors introduced an interactive storytelling framework called SAGE for Storytelling Agent Generation Environment based on an interactive toy. Their storyteller is embodied in a stuffed animal and it is conceived to support two kinds of interactive storytelling framework; the one engaging children with ready-made stories and the one involving children as active story authors (known as Story-making framework). It has been shown that, by interacting with SAGE, children unconsciously explored different aspects of their own identity such as their fears, feelings, interests, and social roles [10]. Furthermore, in [3], the authors focused on the benefits of using virtual peers in storytelling framework with disabled children. In this work, virtual peers have been used as a mean to establish a natural interaction with disabled children (e.g. with Autism Spectrum Disorder) in order to help them to improve their cognitive and communicative functions and skills.

This line of research focus on the development of interactive storytelling frameworks to establish a natural interaction, without considering the expression of affects. However, expressing and acting emotions and affect during the storytelling task is also equally important to deliver a storytelling experience that is as engaging as provided by a human storyteller [34]. Besides, emotion expression can have a considerable impact on the believability of an artificial storyteller [11].

Recently, there has been a growing interest on the development of an expressive storyteller. The need to generate expressive and emotional speech was highlighted and considered for the development of expressive character-based storyteller in [11] and [34]. Other studies focused on bodily expression of emotions and moods during storytelling task. The NAO robot is used in [19] for this purpose. In [32], the authors present their framework of expressive virtual storyteller called Papous. A text-to-speech framework was used in their work and expressive parameters were employed to endow the virtual storyteller with the ability to express emotion through prosody. However, the authors claimed that the agent’s voice still seems as synthetic despite the efforts that have been made to make it sound as expressive.

New databases of expressive storytelling have been recently proposed in previous works [24, 35]. McKeown et al. [24] proposed a new database to study expressive interactive storytelling. The database includes various annotations regarding dyadic social and expressive interactions including turn-taking and facial expressions with a particular focus on laughter behavior. Volkova et al. [35] focus on bodily expression of emotion during narration. They asked actors to imagine they were telling a story to a child and they were free to express emotions related to the story through prosody, facial expressions and body gestures. These databases could be used to model the emotional behavior of artificial storyteller [19]. However, it is not coupled with the supposed child’s behavior during storytelling, which makes it difficult to guarantee that the expressive storytelling framework will be able to adapt to the interlocutor.

In conclusion, a wide range of existing work focuses on the development of interactive storytelling frameworks, in particular to address the educational values of interactive storytelling. However, the modeling of an expressive narration, in particular with virtual peers, has received less attention. Besides, studying the combined impact of emotion expression and joint attention initialization on the experience of Storytelling with children has so far received little attention. In this paper, we focus on the modeling of an expressive narration within a storytelling framework based on a virtual character. Our short-term goal is to study the impact of emotion expression on the appreciation, the engagement and the cognitive processing of children during the storytelling experience.

3. METHODOLOGY

Our expressive virtual storyteller is based on the MARC virtual agent platform, a platform for Multimodal Affective and Reactive Characters [12]. The female agent, named Mary, is used for our study (see Figure 1). We aim to use Mary as an expressive and interactive storyteller with children of 6-8 years-old. 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. Suddenly the ball gets stuck on a roof. They try to get the ball down by throwing their personal items on the roof. However, the situation gets even worse since all the items also remain on the roof. Eventually, the wind blows all the items and the ball off the roof.

Our methodology is based on two-step approach. The first step consists in a manual annotation process where the story content is provided with tags that will serve later to modulate the nonverbal behavior of the agent. The second step refers to a real time animation of our virtual storyteller according to the story content and the appropriate tags of
nonverbal behavior. Since our goal is to endow our virtual storyteller with the ability to drive an interactive and expressive narration, the story content has been revised by researchers in psychology to add interactive dialogues. Such an interactive setting helps to establish an interactive narration task, to stress the emotional content of the story events, and to initialize and maintain the attention of children. The story content has been split into 160 utterances.

160 audio files are used for the 160 utterances. In the present study, we aim to endow our agent with emotions mainly through facial expressions. We do not consider prosody features neither bodily expression. 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. We synchronize lips motion provided from JSAPI text to speech tool with the audio recording of narration. The utterance content is used to drive lips motion.

We distinguished 3 categories of utterances: 1) Joint Attention initialization (JA) utterances, 2) Interactive utterances and 3) Narrative utterances. **JA** utterances (22.50% of utterances) are used to direct the child’s attention to events or objects related to the story (e.g. “look at this teacher”). **Interactive utterances** (21.50% of utterances) mainly refer to a question meant to maintain the child’s engagement (e.g. “do you think he is going to make it?”) or to the introduction and conclusion phases (e.g. “Hello my name is Mary”). They involve the agent as a peer during narration. Both JA and interactive utterances are useful to maintain engagement during the narration task. **JA utterances** are coupled with nonverbal behaviors to direct the child’s attention to a character/event/object. 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 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. We used the following timing for ordering the different modalities during deictic behaviors: gaze behavior starts with the utterance and it is followed by the torso movement or the hand pointing gesture. While torso turning behavior is immediately followed by the rest posture, the pointing gesture holds until the end of the next utterance (as shown in the third image of Figure 1). Finally **Narrative utterances** (56.25% of utterances) correspond to the story content.

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 we displayed them next to the virtual storyteller (see Figure 1). The agent can use deictic behaviors on these images. During narration, a pause of 400ms is made between successive utterances. Without including the interaction of the child during storytelling task, the narration of the story lasts around 10 minutes. During the storytelling task, the expressive storyteller expresses emotions through facial expressions. These emotions are computed from the content of the utterances and they are modeled based on an appraisal theory. In the next section, we illustrate our expressive framework through the evaluation and the annotation of emotion expression according to the narration. We also present the computational model used to endow our virtual agent with emotions.

4. EXPRESSIVE NARRATION

**Emotion annotation and evaluation.**

Different theories were proposed according to the literature in psychology to model the morphological and dynamic characteristics of facial expression of emotions. They are commonly regrouped into three approaches: Categorical, Dimensional and Appraisal approaches. Appraisal theory is one of the most influential theory within affective computing [22] [5]. It has been widely used as the basis for several computational models of emotion [21]. 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 [31]. Five appraisal variables, also called “Stimulus Evaluation Checks” (SEC), have been selected and considered for our study: Expectedness, Pleasantness, Goal Hindrance, Coping Potential and Coping Power [31].

The evaluation of these appraisal variables requires a clear definition of the storyteller’s viewpoint. The same event can be considered as pleasant or not depending on the point of view adopted by the storyteller. An example of such an event considered in the story used for our study is when the ball gets stuck on the school’s roof. While this event may seem surprising and disappointing from the story character point of view, it is from the storyteller’s point of view quite expected (based on the story’s title) and quite funny. In order to provide clear insights into the evaluation of each event, we define three viewpoints the agent can adopt to evaluate the emotional content of the text:

- “Narrator”: In a “narrator” point of view, the agent evaluates narrative, interactive or JA utterances according to his state of mind. In our study, the narrator attitude is defined as friendly, kind and benevolent. The agent also tries to make the child aware of the social norms. We consider only two appraisal variables to evaluate an event from the narrator viewpoint: Expectedness and Pleasantness. Only Expectedness is activated for JA utterances as the main goal is to draw the attention of the child to a particular event or object in the scene. For example, on the second image of Figure 1, the narrator comments the pupil kissing his shoe when he finds it. The act of kissing the shoe is evaluated as unexpected (Expectedness=−0.5) and unpleasant (Unpleasantness=0.5). This image shows a screen-shot of the agent’s facial expression corresponding to this evaluation.

- “Empathic Narrator”: The second point of view is defined as an “Empathic Narrator” where the narrator evaluates narrative and JA utterances. The agent anticipates the emotions according to the characters attitude and the following story events. The resulting evaluation of appraisal variables constitutes a blend of agent interpretation of events and agent anticipation of characters attitude. For instance the narrator is wondering how the children are going to manage the situation as their personal things are stuck on the roof. The present event is not pleasant from the children viewpoint (Unpleasantness=0.5) neither in line with their goal (Goal Hindrance=0.5). Knowing that the “Empathic Narrator” anticipates the events and knows that the classes will start soon, the children cannot control
the situation (Coping Control=-0.5) and they do not have the required resources to deal with it (Coping Power=-0.5). A screen-shot of the agent’s facial expression corresponding to this evaluation is shown in the first image of Figure 1.

• “Story Character”: The third and last point of view is a “Story Character” viewpoint where the storyteller evaluates a given story event according to the character’s situation, that is embodying the emotion of the character. For instance, the child jumps for joy when he finds his ball after the storm (third image on Figure 1). This event is evaluated as unexpected (Expectedness=-0.5), pleasant (Unpleasantness=-1) and in line with his goal (Goal Hindrance=-1). Besides, the child can exert control (Coping Control=0.5) as he can start playing again and he has the required resources (the ball) to do so (Coping Power=0.5).

The appropriate viewpoint and appraisal evaluation is attributed to each utterance according to its content. We also manually correct the evaluation of appraisal variables according to the recorded speech to ensure the coherence between vocal and facial expressions. At the end of the annotation process, only 57.50% of the utterances were evaluated as expressive (i.e. at least one appraisal variable was not set to neutral). They are composed of: 73.91% of narrative utterance, 6.52% interactive utterance, and 19.57% of JA utterance. To evaluate the expressive content of these utterances, the agent adopted respectively a Narrator, a Character and an Empathic narrator viewpoints in 32.61%, 45.65%, and 21.74% of expressive utterances. Indeed, emotions were mostly evaluated according to the story’s content and the characters’ attitudes.

Computational model of facial expression.

The mapping between the appraisal variables and the facial actions units as proposed by the FACS (Facial Action Coding System) has been clearly established in previous works [31]. This allows us to build the related computational model of facial expressions. For instance, a surprising event (low value of Expectedness check) mainly implies the following Action Units: 1 (Inner brow raiser), 2 (Outer brow raiser), 5 (Upper lid raiser) and 26 (Jaw drop).

A computational model of these five appraisal variables has been developed within the MARC platform. The mapping between each appraisal variable and the associated facial action units is mainly based on the work described in [31]. A continuous value that belongs to the interval [-1 1] can be attributed to each appraisal variable. In order to simplify the annotation of appraisal variables, we use discrete values defined along a 5-points scale for Pleasantness and Goal Hindrance (from -1 to 1). However, we only use a 3-point scale (-0.5; 0; 0.5) for Coping Control and Coping Power variables as the facial actions units related to a low control and power imply a strongly negative facial expression (a strong lip corner depression) which may considerably disturb children during narrative process. Expectedness is also evaluated according to this 3-point scale. Indeed, the dynamic process of facial expression related to the lowest amount of Expectedness (i.e. -1) implies a high amount of eyebrows raise and of mouth open, which does not infer a natural behavior during narration.

Temporal relationship between emotion and speech.

The coordination between verbal and nonverbal behaviors of a virtual character may have a considerable impact on the believability of a virtual character [23]. Two dimensions were particularly studied in our expressive storyteller: when to express the emotion and how to handle the neutral expression.

Previous works showed that different configurations of the temporal relationship between facial expressions and speech utterances may have a considerable impact on the recognition, the perceived intensity and the realism of facial expressions [8]. Based on the results of a perceptual experiment, it has been shown in [8] that facial expression of emotions displayed on a virtual character’s face seems to be perceived as more intense and more realistic when displayed during the whole utterance. In our work, facial expressions of our virtual agent are displayed during the whole utterance to increase their intensity and their realism.

Narrative is composed of expressive and non-expressive utterances. During expressive utterances, the virtual agent display the corresponding facial expression according to the annotation of appraisal variables. During non-expressive utterances, the virtual agent is not supposed to display a particular facial expression but a neutral face. However, previous works have discussed the theoretical concept and the qualitative feasibility of neutral faces during natural interaction [9]. Neutral faces can be also perceived as inferring negative emotional states [9]. Endowing a virtual agent with smiles during an interaction can create a friendly atmosphere, which is of high importance during an interaction with children. Besides, a smiling face matches better the...
profile our virtual storyteller defined as friendly and benevolent. According to their social function, different types of smiles were identified in previous studies such as polite, embarrassed and amused smiles [28]. Amused smile was considered as the most appropriate to be adopted by a benevolent and friendly agent. Based on Facial Action Coding System (FACS), some insights were highlighted in [28] to define the morphological characteristics of an amused smile. For instance, facial expression of an amused smile mainly involve the following Action Units (AU): 6 (Cheek raiser), 12 (Lip corner puller). In addition to these Action Units, we also use AU1 (Inner brow raiser) to slightly raise the eyebrows as the Neutral eyebrows configuration of our agent seems odd when coupled with a smile. AU25 (lips parted) is also used to alternate between two categories of smile during narration. Indeed, one of these two smiles are randomly displayed for each non-expressive utterance. During the 400ms pauses between utterances, the facial expression displayed by the agent correspond to the one attributed to the previous utterance.

5. CONCLUSIONS AND FUTURE DIRECTION

Our expressive storyteller architecture supports the expression of emotions according to the story content but also according to the point of view of the storyteller. We highlighted three different viewpoints that the storyteller can adopt to evaluate a story event: Narrator, Empathic Narrator and Story Character. We use the Component Process Model appraisal theory to evaluate the expressive content of the utterances during narration as a continuous changes in the states of different emotion functions in response to the evaluation of an event. This drives the agent’s Action Units to express emotions, depending on the point of view of the storyteller.

In short-term future work, we aim to study the effect of emotion expression on the appreciation, the cognition and the engagement of children during storytelling experience. To achieve these goals, we aim to conduct a perceptual experiment with two groups of children of 6-8 years old. The first group of children will be asked to participate to the storytelling task driven by our expressive virtual storyteller. The second group will be engaged in the same storytelling task driven by our non-expressive virtual storyteller. A within subject design will be adopted to study the relevance of emotion expression during storytelling task with a virtual peer. Firstly, we assume that endowing the narrative virtual agent with emotions during narration enhances the ability of the child to memorize the content of the story. Finally, we assume that endowing the narrative virtual agent with emotions during narration task would make the child more engaged and reactive. An EyeTracker system will be used to track the direction of the child’s gaze allowing the quantification of some engagement measures such as the frequency and the duration of the child’s gaze toward the agent face and toward the narration screen.

In the long-term future work, we also aim to enrich the nonverbal communication of emotions by including emotional body expression. In the present work, the evaluation of the appraisal variables as well as the choice of the storyteller viewpoint have been performed based on a manual annotation. The automatic evaluation of these components according to the story content and to the agent memorization of event could be of high interest for the development of an autonomous expressive narration computational model. Our current interactive framework is based on a “Wizard of Oz” approach. The virtual storyteller can listen to the answer of the child after interrogative utterances. However, handling the interruptions made by the child during the storytelling task will enhance the interaction quality.

6. ACKNOWLEDGMENTS

The authors would like to thank Matthieu Courgeon for his help in the development of our virtual storyteller. Part of this work was funded by the Agence Nationale de la Recherche (ANR), project NARECA (ANR-13-CORD-0015).

7. REFERENCES


