Tell Me That Bit Again... Bringing Interactivity to a Virtual Storyteller

André Silva[1], Guilherme Raimundo[1], and Ana Paiva[1]

IST and INESC-ID, Rua Alves Redol 9, 1000 Lisboa, Portugal
{andre.silva,guilherme.raimundo,ana.paiva}@gaips.inesc.pt

Abstract. Stories and storytelling are a constant presence in our lives since very early childhood. Who does not remember a story narrated by a good storyteller? However, real human storytellers do not always tell the story the same way. They observe their "audience" and adapt the way they are telling the story to better respond to their reactions.
This paper focuses on how to bring interactivity to a virtual storyteller by allowing users to influence the story. The storyteller is a synthetic 3D granddad that uses voice, gestures and facial expressions to convey the story content to be told. The character's behaviour and the way the story is narrated, is influenced by the user's input. Such input is done by a tangible interface (a kind of mail box) where children put the cards they want in order to influence what will happen in the story being told. A preliminary usability test was made with sixteen children, with ages between nine and ten years old. The results showed that the way interactivity is introduced was quite successful.

1 Introduction

We are all storytellers. Stories and storytelling are a constant presence in our lives since very early childhood. Children like to be told a story, over and over again, every time enjoining the situations, the sounds, the events and the worlds portrayed in the story. But, it is not only the story that matters.

The storyteller, himself, plays a fundamental role in children's stories, dragging them into the story, keeping their attention and freeing their imagination. In fact, a storyteller can turn a story into a good or a bad one. The use of the voice, facial expressions, and the appropriate gestures, are basic ingredients for transforming the content of a simple story into the most fascinating narrative we have ever heard. But this need for a storyteller to be expressive, to balance the words and the tone, to use gestures appropriately, poses major research challenges if one aims at building a "synthetic" storyteller. However, recent developments of embodied agents such as [3], [2], [7] and [9] among others, have shown amazing advances, which allows us to consider the technical challenges for building a virtual storyteller can in fact be overcome and achieve, under limited circumstance, a believable storyteller.

In fact, in [11] a simple virtual storyteller was already presented. Our ultimate goal is for the virtual storyteller to be able to tell the content of a story in a natural way, expressing the proper emotional state as the story progresses and capturing the user's attention in the same way a human storyteller would.

O. Balet et al. (Eds.): ICVS 2003, LNCS 2897, pp. 146–154, 2003.

But, storytelling is not only narrating a text in a compelling way. It also involves understanding the audience, reacting to it and even adapting the story and the way the story is told to the cues given by the audience.

In the work here presented, we will show how we have incorporated interactivity in a virtual storyteller, which will adapt certain aspects of the story being told to some "cues" the user/child will be providing during a storytelling session.

This paper is organised as follows. First we will describe the idea for the interactivity of the storyteller. Then we describe the character, the structure and contents of the stories embedded in knowledge of the character. Then we show how the user influences the stories being told, and draw some final conclusions.

2 The Idea

Real human storytellers do not always tell the story the same way. They observe their "audience" and adapt the way they are telling the story to better respond to their reactions. This means that the storyteller gets feedback from his audience and uses that feedback to shape the story the way he or she believes it should be told, at that particular moment. Using this idea our virtual storyteller must be interactive, adapting the story being told to the input he gets from the user. So, the storytelling process will follow a behaviour as shown in Figure 1.

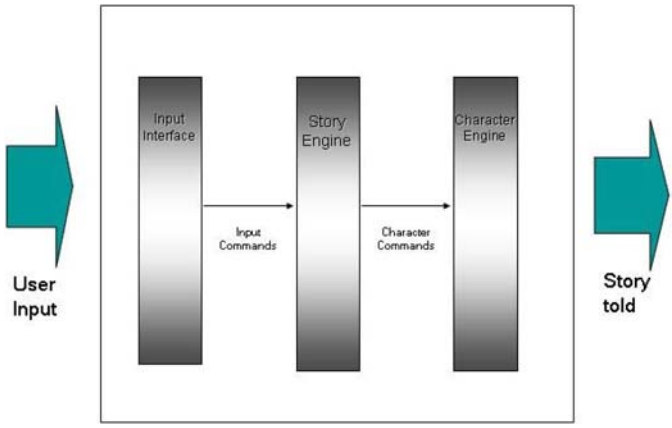

Fig. 1. System Behaviour

The user supplies the virtual storyteller with certain input, which can be done in several different ways, allowing the character to decide how the story should be told. For instance, the user may decide he wants to hear a more terrifying version of the story, supplying this simple wish to the virtual storyteller. The virtual storyteller is then responsible for choosing the course of the story that is most suitable for the user's input and for adapting his visible behaviour to the user's intentions. The Input Interface

module is responsible for receiving and handling the input from the user, organizing it for future processing by the story engine. Also, since the input can be supplied at any time, it is also this module's responsibility to store it in a coherent way when the Story Engine module isn't ready to process it yet.

The Story Engine module contains the story itself, parsed from a story file. This module is responsible for parsing the story, organizing it and maintaining the necessary information to decide how to tell it, according to the input received from the Input Interface module. Note that here, a story is intended to be a non-linear story. Depending on the decision made by the Story Engine, different commands are sent to the Character Engine. The Character Engine is responsible for processing the bits of story that need to be told guaranteeing that the synthetic character performs the adequate actions, moves and gestures to convey the desired meeting. Moreover, the Story Engine tells the Character Engine how it should set the character's behaviour in order to maintain coherence with the direction that the story will take.

3 The Character

The storyteller is a synthetic 3D granddad that uses voice, gestures and facial expressions to convey the story content to be told. The character's behaviour and the way the story is narrated, is influenced by the user's input. For instance, the user may express that he or she would want the story to be told in a more scary way. Therefore, the character's emotions are affected by the user's choices. Consequently the virtual storyteller's visible behaviour (its facial expression, voice and gestures) is also influenced by the user's input. The character's verbal output is affected gradually, allowing for several levels of emotional change in the voice. This is done by changing the Text-to-Speech system's parameters (the Eloquent TTS system), adjusting them so they convey the emotion the character is trying to express. For instance, the character can be mildly sad or very sad. The same concept was applied to the facial expression of the virtual storyteller. The facial expression engine follows the MPEG-4 standard [8] in which the six universal emotions of joy, sadness, angriness, disgust, surprise and fear [4] are contemplated. According to the present emotional state of the character the emotional facial displays are blended together. This way it is possible to convey several emotions simultaneously having each contribute to the final output with a specified weight.

At the moment the lip-synch of the model is still very simple. When the voice is heard the facial engine generates random visemes (the visual equivalent of a phoneme). Due to this random nature the visual output you get is cartoon-like where the facial display doesn't match exactly the audio. However, due to a cartoon-like appearance of Papous, although quite simple, this approach, still widely used, leads to quite satisfactory results.

4 The Stories

A human author creates the story files for our synthetic character to narrate. The stories themselves are written using a specific interface and are represented in the story

file in XML format. This story file is initially parsed and organized by the Story Engine module until it becomes ready to be narrated. Figure 2 depicts the story creation interface.

Fig. 2. Story Creation Interface

The story is created in levels. There can be as many levels as wanted. When the virtual storyteller progresses between different levels, he is progressing towards the story's end. In each level, there are multiple StoryBits (inspired by the notion of Beat [6]), each one with its different properties (each has a function according to Propp's [10] functions), characters and events. The author of the story defines not only the several StoryBits in each level but also as many levels as he sees fit, setting the different properties for each one of them separately. Each bit can be connected to many different bits in the next level. Thus, the author can make his story as flexible or as linear as he or she wants. Figure 3 depicts an example of a possible story structure, with several levels and various StoryBits for each level.

If the author wants a very flexible story, with lots of different ways the story can progress, s/he can choose to have many levels and several StoryBits for each level. If, on the contrary, the author of the story wants a linear narrative, all s/he has to do is to define only one StoryBit available for each progression level. Note that in the example shown above, the author has decided to create a single StoryBit in level 3, which means that all the story instances will have at least that particular stage in common. On the other hand, the story can have two different ways it can start, depending on the input that the user decides to provide before the story commences (see the next section for details).

Also, the author creates the Storybits in each level separately, which means that he cannot create Storybits from different levels at the same time. Thus, the author is encouraged to create each stage of the story individually, defining for that particular level, all the possibilities that may occur. When he is finished defining the level, he may lock it (see figure Figure 2) and proceed to the next level. This is done to facilitate the creation of non-linear stories which can become a very tiresome task (trying to maintain a clear idea on what happens in each parallel "story reality").

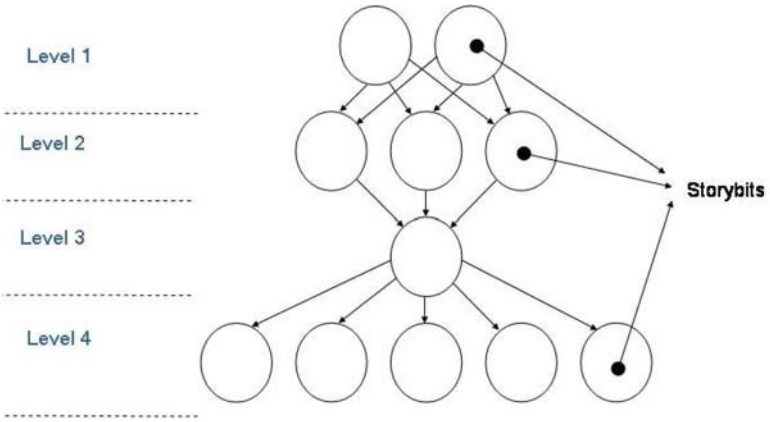

Fig. 3. Basic story structure

5 Interactivity

When the virtual storyteller progresses through the levels, it must choose which Story-Bit to narrate, according to the user's input. There are two underlying problems to solve here: (1) the first one is the user's input. What type of input can we get from the user? The second one is navigation. That is, given the user's input, how will the storyteller decide which bit to pick next, maintaining the coherence of the story?

5.1 User's Input

Concerning the first question, we decided to use a tangible interface that will get the user physically involved with the story. We investigated the use of several different types of input, such as voice or even SenToy [1]. However, due to the fact that we wanted to provide some "story meaning" to the input, and at the same time get the user physically involved, we decided to use the Influencing box [5], as we could associate "images" and meanings to the input. With the box, the user just needs to insert illusive cards (which are tagged with bar codes). The user may choose the most appropriate card for a particular situation (for example, choose a scary sign or a forest) which will then influence the whole story. Figure 4, depicts the use of the Influencing box as input for the system.

The user stands in front of the box, which allows him to insert the cards without much effort, while the character itself is projected onto a wall. Before the story commences, the user can insert cards too, and this input information will be used to decide how the story will start, a kind of setting up the scene (this "pre-story" input will only be considered by the system if the human author has decided to provide more than one way for the story to begin).

Inside the Influencing box, these cards are identified and their meaning is sent to the Input Interface module for processing (shown previously in Figure 1).

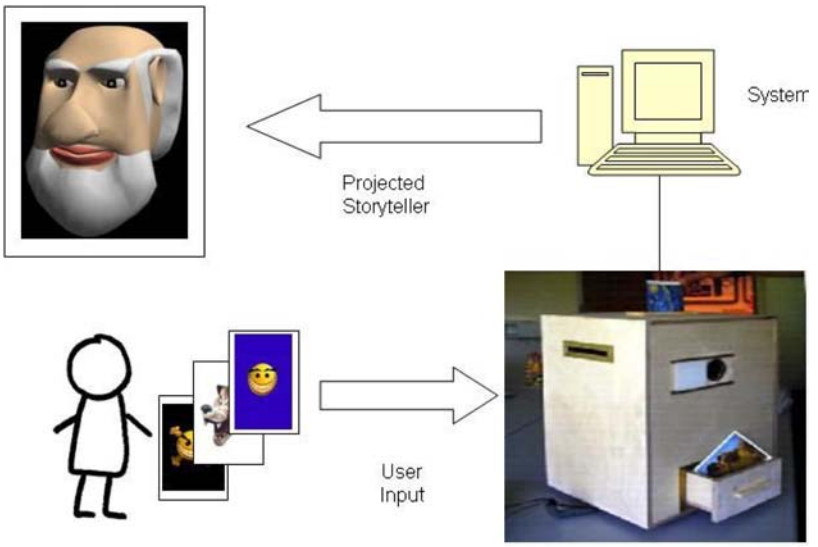

Fig. 4. Input using the Influence Box

There are four types of cards: Propp function card's; character cards; mood cards and scene cards. All the cards are available all the time, which provides a large diversity on the possible progressions in the story. The user can supply these cards at any time, and their meaning is stored by the system until it becomes ready to use that information, which means, when it becomes necessary to choose the next Storybit for narration.

5.2 Navigation through the Story

The input supplied by the user is used to guide the virtual storyteller, helping him decide which StoryBits to choose. Thus, according to the user's input, the storyteller decides which of the different StoryBits available at the current level he should choose to narrate.

Although the stories' are quite simple in terms of non-linearity, given the knowledge structure associated to each StoryBit, the process of navigating through StoryBits is quite rich and is based on heuristics that rely on the StoryBits properties (the emotion/mood of that bit, the characters, the scenario and the Propp function of the bit).

The decision made by the storyteller is based on a desirability factor that all StoryBits have. Naturally, this desirability factor is not a constant and depends not only on the input received at each moment as a card but also from the previously narrated StoryBits.

Thus, with each level transition, the virtual storyteller calculates the desirability factor for each available StoryBit and then decides which one to choose. The system calculates the desirability factor by trying to match as many available StoryBit properties as it can with the user's input (the chosen StoryBit is the one with the highest desirability factor). For instance, if the user wishes the story to be told in a happier way,

then the virtual storyteller will try to accommodate his wish by choosing the appropriate StoryBits in the future.

The user inputs themselves are received and weighed, considering the time when they arrive and their type. This means that the most recent input has a bigger part in the decision that the storyteller has to face with each level transition.

For instance, let us assume the user decides that he wants the story to be told in a scarier way and provides the necessary input. He can then change his mind and decide he wants the story to be sadder. The user's most recent input is more important for the choice of which StoryBit to narrate and therefore, the next StoryBit will be coherent with his later choice, making the virtual storyteller tell the story in a sadder way.

However, there are two situations that may occur, with which the system has to be capable of dealing with: the lack of input provided by the user and the lack of adequacy between the user's wishes and the StoryBits available for narration.

If the user chooses not to provide input to the system, then the virtual storyteller assumes that the story is pleasing the user and, in the next level, the character searches for StoryBits as similar as possible to the ones he has already narrated. This way, if the user is enjoying the way the story is being told, the system maintains coherence throughout the rest of the story. This system behaviour is obviously consistent with a real human storyteller's actions.

When the user provides input that cannot be satisfied by the StoryBits available for narration, the system uses a similar approach. This means that, unable to suit the user's intentions, the system chooses the next input as if there had been no input from the user, i.e., choosing the next StoryBit according the one previously narrated. This approach seems reasonable, considering what a real human storyteller would do in this situation, and it yields quite satisfactory results.

6 Results and Final Comments

A preliminary usability test was made with sixteen children, with ages between nine and ten years old, in order to evaluate the text to speech engine, the facial model and facial expressions and the tangible interface when comparing to a classic menu. The test consisted of the narration of the story of Little Red Ridding Hood and a questionnaire that children answered after hearing and interacting with the virtual storyteller.

Based on the given answers we can draw the following initial conclusions:

- The application is of easy use
- The main obstacle to the understanding of the story is the quality of the text-to-speech
- The "looks" of the story teller were pleasing
- Sometimes, the facial expressions were not completely identified by the children
- The tangible interface was a success having 14 children chosen this one over the classic menu interface

The story teller is still at an early stage having several aspects that need to be improved. After our analysis we think that using another text-to-speech or human recorded

voice provided by a professional actor will improve considerably the quality of the overall application. However this last solution poses some problems because we want to use a continuum of emotional state. Even if we consider discrete intervals, the number of recorded sets needed to achieve a smooth transition might not be feasible. Regarding the model we concluded that it needs to be more cartoon-like and that the facial expressions need to be more exaggerated.

Another aspect that should be improved is the story creation interface. Future work will include the facilitation of the interactive stories' creation. In fact, creating a non-linear interactive story with the current version of the system's Story Creation Interface is a tiresome and difficult task, although feasible. Trying to maintain a clear idea on all that is happening in all of the possible versions of the story is a complicated job for a human being, which should be left for the computer to handle.

Thus, including more information regarding the narrative context dependencies of all the possible versions of the story would allow the computer to become more effective in assisting the creation of the non-linear interactive stories.

The final version of the storyteller's body is still being improved, but the character's gestures and body expression will also be influenced by the emotional state of the character [12]. By connecting the input from the user with the utterance of the voice, the emotional state, facial and body expression in an interactive emotional storytelling system, we hope to deliver a pleasing and didactic experience to the user.

Acknowledgements

Thanks to Fernando Rebelo for the design of Papous. Thanks to the GAIPS team for their comments and criticisms during the development of this system. Thanks to Phoebe Sengers for allowing us to use the Influencing Machine.

The work on Papous was funded under the Sapiens Program- Fundação para a Ciência e Tecnologia, project number POSI/SRI/41071/2001.

References

1. P. A., A. G., H. K., M. D., C. M., and M. C. Sentoy in fantasya: Designing an affective sympathetic interface to a computer game. *Personal and Ubiquitous Computing Journal (Ub Comp)*, (5-6), 2002.
2. J. Cassell, T. Bickmore, L. Campbell, H. Vilhjalmsson, and H. Yan. Conversation as a system framework: Designing embodied conversational agents. In e. a. J. Cassell, editor, *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 1999.
3. J. Cassell. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In J. C. et al., editor, *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 1999.
4. P. Ekman. *Emotion in the Face*. New York, Cambridge University Press, 1982.
5. S. et al. The enigmatics of affect. In *Proceedings of Designing Interactive Systems-DIS'2002*. ACM Press, 2002.
6. M. Matheas and A. Stern. Towards integrating plot and character for interactive drama. In *Socially Intelligent Agents: Creating Relationships with Computers and Robots*. Kluwer, 2002.

7. T. Noma, L. Zhao, and N. Badler. Design of a virtual human presenter. *IEEE Computer Graphics and Applications*, 20(4):79–85, 2000.
8. I. Pandzic and F. R. (editors). *Mpeg-4 Facial Animation - The standard, Implementation and Applications*. John Wiley & Sons, 2002.
9. C. Pelachaud and I. Poggi. Subtleties of facial expressions in embodied agents. *Journal of Visualization and Computer Animation*, 2002.
10. V. Propp. *Morphology of the folktale*. Austin: University of Texas Press, 1968.
11. A. Silva, M. Vala, and P. A. Papous: The virtual storyteller. In *Intelligent Virtual Agents*. Springer, 2001.
12. M. Vala, M. Gomes, and A. Paiva. Affective bodies for affective interactions. In *Animating Expressive Characters for Social Interactions*. John Benjamins Publishing Company, 2003.