To tell or not to tell...Building an interactive virtual storyteller

Andre Silva*
*INESC-ID and IST, Tagus Park Av. Prof. Cavaco Silva, 2780-990 Porto Salvo, Portugal
andre.silva@gaips.inesc.pt

Guilherme Raimundo
†INESC-ID and IST, Tagus Park Av. Prof. Cavaco Silva, 2780-990 Porto Salvo, Portugal
guilherme.raimundo@gaips.inesc.pt

Celso de Melo
‡INESC-ID and IST, Tagus Park Av. Prof. Cavaco Silva, 2780-990 Porto Salvo, Portugal
celso.melo@gaips.inesc.pt

Ana Paiva
§INESC-ID and IST, Tagus Park Av. Prof. Cavaco Silva, 2780-990 Porto Salvo, Portugal
amp@inesc-id.pt

Abstract

Storytellers do not always tell the story the same way. They observe their ”audience”, see their reactions and adapt the way, the gestures, the posture and the content of the story, to better respond to the audience's reactions. Clearly, this adaptation is however not only in the content of the story but also on the way the story is told, thus the facial expressions and the gestures of the storyteller. In this paper we describe a synthetic character that tells stories interactively, discussing how the adaptation to the audience is done through the content of the story and the associated expressions and gestures of the character.

1 Introduction

Any storyteller 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 [2], [3], [6] and [8] 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 circumstances, a believable storyteller.

In fact, in [10] and recently in [9] a simple virtual storyteller was presented and its interactivity discussed. 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.

In the work here presented, we will show how we have adapted the storytelling (both content, gestures and expressions) to the interactivity in a virtual storyteller.

This paper is organised as follows. First we will describe the idea for the storyteller. Then we describe the character, the structure and contents of the stories embedded in knowledge and the gestures of the character. Then we describe how the user will influence the stories being told, and the gestures of the character.

2 Papous, the storyteller: 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, their expressions and gestures to better respond to the audience's reactions. This means that the storyteller gets feedback from his audience and uses that feedback to shape the way in which the story is delivered.

So, our main research question is: how can we adapt the gestures and expressions of the character to be able to adequately respond to the audience.

To do that, we have built an interactive storyteller that adapts to the user's input using a tangible interface as shown in Figure 1.

The user supplies the virtual storyteller with certain input (in this case ”postcards”, which will allow 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 storyteller gestures and expressions change in order to look frightened and scared in order to create the right response from the audience.

To do that, the architecture presented contains a set of modules, in particular:

- *(1) the Input Interface module* is the component re-

sponsible for receiving the input from the user and then handling it by 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.

- *(2) 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.

- *(3) The Character Engine* is the component that handles all the processing of the story bits that need to be told, thus guaranteeing that the synthetic character performs the adequate actions, moves and gestures to convey the desired meaning. 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.

2.1 The Character

The storyteller is a synthetic 3D granddad (inspired in a set of TV granddads such as the Tweenies and "Avô Cantigas" (in Portugal). The character takes advantage of its 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.

The facial expression engine follows the MPEG-4 standard [7] 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. In 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.

3 Interactivity

As the virtual storyteller progresses through the levels, it must choose which StoryBit 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?

3.1 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 1, depicts the use of the Influencing box as input for the system.

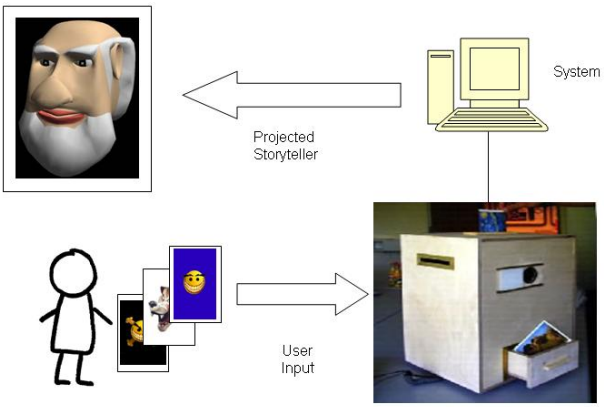

Figure 1: Input using the Influence Box

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).

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.

3.2 Adaptation to the Audience

The input supplied by the user is used to guide the virtual storyteller, helping him decide which StoryBits to choose and the gestures and expressions to make.

Although the stories are quite simple in terms of non-linearity, given the knowledge structure associated to each StoryBit (which contains basically a small piece of story), 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). All this process of deciding what to tell next and how to tell 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, taking into account the time when in which 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.

3.3 Adaptation of the storyteller's gestures and expressions

Emotions are intangible, intimate and personal. They are also the main elements that need to be changed in the course of a narration. However, to express emotions, one requires the adequate and coordinated use of vocal, facial and body expressions. Furthermore, to express emotions in a manner that can be easily perceived by the audience, requires an understanding that artists in Dramatic Expression and Animation seem to dominate. Thus, a good storyteller also needs to express himself dramatically. So, if user input affects Papous' internal emotional model, then one important question is: How will emotional state change reflect on its expression? If the storyteller is sad its gestures should be slow and narrow. However, if it is enthusiastic its gestures should be energetic and wide.

Our approach to this problem relies on two main aspects: (1) high-level parameters that define movement qualities; (2) A theory that correlates emotions to these parameters.

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.

4 Results and Final Comments

Although still at at the beginning, 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 compared with a classic menu. The test consisted on 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 were able to 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 nice
- Sometimes, the facial expressions were not completely identified by the children
- The tangible interface was a success (fourteen out of sixteen children chose it over the classic menu interface)

As the story teller is still at an early stage, there are obviously several aspects that need to be improved. In particular, the TTS, which we need to consider also the use of a human recorded voice provided by a professional actor.

Secondly, as the final version of the storyteller's body is still being improved, the character's gestures and body

expression will be then more developed. 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.

5 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- Funda cão 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. 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.

[3] 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.

[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] T. Noma, L. Zhao, and N. Badler. Design of a virtual human presenter. *IEEE Computer Graphics and Applications*, 20(4):79–85, 2000.

[7] I. Pandzic and F. R. (editors). *Mpeg-4 Facial Animation - The standard, Implementation and Applications*. John Wiley & Sons, 2002.

[8] C. Pelachaud and I. Poggi. Subtleties of facial expressions in embodied agents. *Journal of Visualization and Computer Animation*, 2002.

[9] A. Silva, G. Raimundo, and A. Paiva. Tell me that bit again... bringing interactivity to a virtual storyteller. In *Virtual Storytelling: Using Virtual Reality Technologies for Storytelling*. Springer, 2004.

[10] A. Silva, M. Vala, and P. A. Papous: The virtual storyteller. In *Intelligent Virtual Agents*. Springer, 2001.