The Storyteller:
Building a Synthetic Character That Tells Stories

André Silva
IST/INESC-ID
Rua Alves Redol, 9
P-1000-029 Lisbon, Portugal
+351 21 310 03 00
andre.silva@gaiva.inesc.pt

Marco Vala
IST/INESC-ID
Rua Alves Redol, 9
P-1000-029 Lisbon, Portugal
+351 21 310 03 00
marco.vala@gaiva.inesc.pt

Ana Paiva
IST/INESC-ID
Rua Alves Redol, 9
P-1000-029 Lisbon, Portugal
+351 21 310 03 00
ana.paiva@inesc.pt

ABSTRACT

This paper introduces the Storyteller, a synthetic character that tells stories. Our goal is to obtain an expressive and believable embodied agent that can resemble a real human storyteller.

The several aspects concerning the conception of the character, as well as some preliminary results, are presented in this document.

Keywords

Synthetic Character, Gestures, Emotional Expression

1. INTRODUCTION

Stories and story telling remain in our memory since early childhood. Who doesn't remember the stories we were told when we were young? The enchantment of such stories comes, in part, from the presence of the storyteller. A storyteller, more than just reading a text, uses the voice, the facial expression, and appropriate gesticulations and posture in order to convey the ambience and the content of the story.

Aiming at this experience and in line with current developments of embodied agents (see [2], [4], [5], [6] and [7]) we have created the Storyteller. The Storyteller is a synthetic character, immersed in a 3D virtual world, which will narrate the content of a story in a natural way, expressing the proper emotional state as the story progresses.

The work present in this paper is still in an early stage. At the present time, the Storyteller simply acts as a virtual narrator who reads a text enriched with control tags. Such tags allow the storywriter to control the emotional state of the character (thus influencing his behaviour and voice) and order explicit gestures. This approach is similar to the one used by Allison Druin (see [8]) where children annotated text so that a robot could produce the appropriate emotions when the story was narrated.

Additionally, the storywriter can also use tags to control the surrounding environment (set and illumination). The idea is to move the storyteller through the different places mentioned in the story and, that way, achieve a correlation between the story and the ambience.

2. ARCHITECURE

The architecture of the Storyteller is loosely based on the REA architecture described in [6].

It has four essential modules: the Input Manager (IM), the Deliberative Module (DM), the Set and Illumination Manager (SIM) and the Action Scheduler (AS).

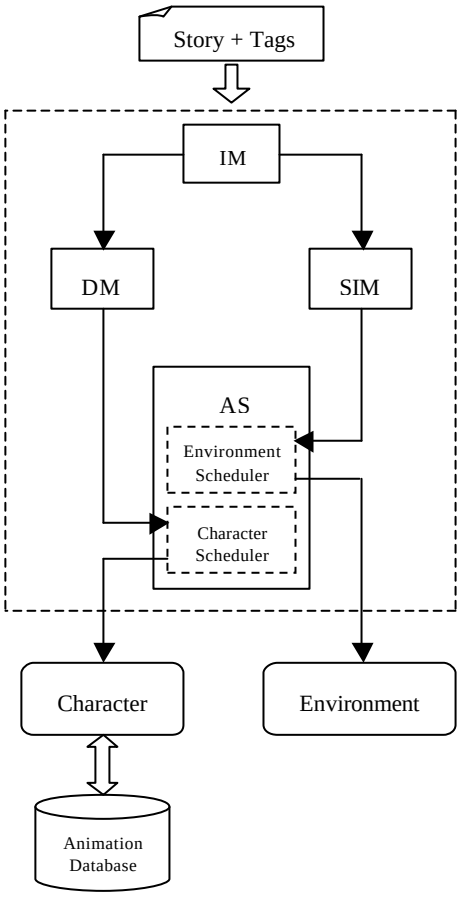

Figure 1 - The Storyteller's architecture

2.1 Input Manager

The Input Manager is the module responsible for processing the text file that contains the story, checking it for syntax and semantic errors, and taking the necessary actions to ensure that the data is correct and ready for the other modules to process.

As input, it takes the annotated story file and a set of configuration files to determine which data values are acceptable for the several kinds of tags available to the user.

The generated output is tag-oriented information: tags with control parameters and the associated piece of text (if any).

2.2 Deliberative Module

The Deliberative Module acts as the Storyteller's mind and, therefore, contains its internal emotional state.

It starts by receiving the configuration information from the Input Manager to initialise the emotional state. Then, it begins to handle its specific tags, namely *emotion* and *behaviour* tags.

The *emotion* tags update the internal emotional state. They indicate which emotion should be changed and the new value that it must have. The new emotional state will affect the character's voice and visible behaviour.

The *behaviour* tags will trigger requests to the Action Scheduler. The requests will result in character's actions such as the generation of speech or the gestures.

2.3 Set and Illumination Manager

The Set and Illumination Manager is responsible for managing the environment where the character is immersed. The environment surrounding the character includes the set and the illumination pattern.

It starts by receiving the configuration information from the Input Manager to initialise the sets and the illumination patterns. Then, it begins to handle its specific tags, namely *set* and *illumination* tags.

The *set* and *illumination* tags update the environmental state, and will trigger requests to the Action Scheduler to change the set and the illumination surrounding the character.

2.4 Action Scheduler

The Action Scheduler is responsible for the generation of the audible and visible part of the application: the character's voice and appearance, as well as the surrounding environment.

As an example, if the Deliberative Module is currently processing a sentence, then it will request the Action Scheduler to synthesize it using the character's voice.

Thus, the Action Scheduler is an abstraction layer between the DM / SIM modules and the external modules: the character engine (the graphics and the text-to-speech systems) and the environment engine.

3. TAGS

We have seen that the text can be annotated with several kinds of tags that affect the character and the surrounding environment.

The Table 1 summarizes the four types of tags and explains the function of each one.

Table 1 – Tag types

Tag Type	Function
Behaviour (1)	Indicates an action that the character should perform (e.g. <**1*big**>)
Set (2)	Specifies a new set where the character should be integrated (e.g. <**2*house**>)
Illumination (3)	Specifies a new illumination pattern (e.g. <**3*day**>)
Emotion (4)	Explicitly modifies the character's emotional state (e.g. <**4*happiness*80**>)

The list of tags is defined in a configuration file and depends solely on the available character's animations and sets. The storywriter is free to use the tags as he pleases, but it should take in consideration the context of the story.

For example, if the writer wants to stress out a particularly scary part of the story, he should specify the appropriate emotional state. The chosen emotional state will change the character's voice and behaviour and, therefore, suit the writer's intentions.

We have defined a very small set of tags for demonstration purposes. The Figure 2 illustrates a possible use of these tags.

<**2*house**> <**4*happiness*80**> Hello everybody! I am extremely happy today! Lets take the usual tour, ok? <**4*happiness*50**> This is the house I live in. It is a very <**1*big**>big house<**~1*big**>.

Want to go outside? <**2*street**> Ahhh.... isn't this nice? I live in a <**1*small**>small town<**~1*small**> right by the sea... Hmmm....I think it will be dark soon... <**3*night**> <**4*fear*80**>Oh, I'm so afraid of the dark... Maybe we should get back in the house, right?

<**2*house**> Hey! Do you like stories? I bet you do! You know, I have a friend named Alex. He is a writer, and he is <**1*tall**>very tall <**~1*tall**>. Much taller than me!... <**4*happiness*20**> I haven't seen him in a while... and that makes me kind of sad...

<**4*happiness*50**> Anyway, I also have <**1*short**>a very short friend <**~1*short**> named Paul. Hey! We have been talking for a long time... It is almost morning! <**3*day**>

<**4*surprise*90**>What a marvellous day! <**4*happiness*50**> Come back soon, ok? Bye,bye...

Figure 2 - Example of annotated text

4. CHARACTER

The character is perhaps the most important element in the story telling scene. We decided to have a 3D character that would tell the story on top of the "told scene", similar to the work by Badler et al. that used a synthetic character as a virtual human presenter [2]. This virtual presenter accepts speech texts with embedded commands as input and performs a variety of appropriate gestures, as a mean to expressively transmit the content of a presentation, to a human audience, in a believable way.

4.1 Emotions

The character's emotional state is internally represented by a set of numerical discrete values. Figure 3 indicates the thresholds established for the three emotions used.

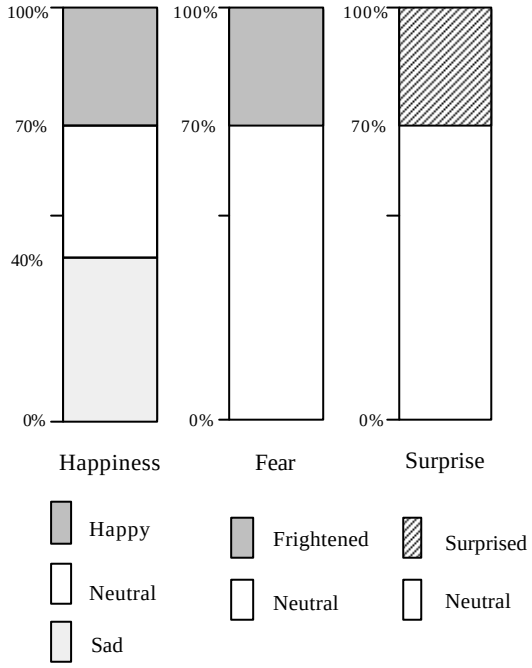

Figure 3 - Emotion thresholds

Of course, other emotions and thresholds could be defined, and further divisions could be considered to have a more refined control over the emotionally affected output properties.

4.2 Voice

The character's voice is extremely important. The precision in the control of the voice depends, mostly, on the text-to-speech (TTS) system.

4.2.1 Voice Parameters

We used a TTS system that allows the control of seven parameters to completely define a character's voice. These parameters are explained in Table 2.

Table 2 – Voice parameters

Parameter	Description
Pitch baseline	Controls the overall pitch of the voice; high pitches are associated with women, and low pitches with men.
Head size	Controls the deepness of the voice.
Roughness	Controls the roughness of the voice.
Breathiness	Controls the breathiness of the voice; the maximum value yields a whisper.
Pitch fluctuation	Controls the degree of fluctuation of the voice.
Speed	Controls the number of words spoken per minute.
Volume	Controls the volume of the voice, i.e., how loud it sounds.

To transmit emotions through the character's voice we established a series of relations between emotions and voice parameters using theoretical knowledge about the way emotions and speech affect each other [10]. Table 3 indicates which parameters should be changed in order to transmit the emotion we intend with the voice.

Table 3 – Emotions / TTS parameter correlation

Emotion	Parameter	Action
Happiness/Sadness	Speed	Increase/Decrease
	Pitch Baseline	Increase/Decrease
	Pitch Fluctuation	Increase/Decrease
Fear	Pitch Baseline	Decrease
	Pitch Fluctuation	Increase
	Breathiness	Increase
Surprise	Pitch Baseline	Increase
	Pitch Fluctuation	Increase
	Speed	Decrease

4.2.2 Pauses and Text Punctuations

Pauses are very important to achieve natural speech. It is of critical importance that the character's voice pauses appropriately, considering the punctuation marks used.

To achieve the correct treatment of text pauses we classify the text into categories, which allow the application to be more specific in processing and interpreting the different punctuations marks that appear in the input file. Table 4 explains the different categories and the pause length associated with each of them.

Table 4 – Text pauses

Text Category	Form*	Pause (ms)
Words	<Sentence>	100
Exclamation	<Sentence> !	500
Interrogation	<Sentence> ?	200
Period	<Sentence> .	200
Comma	<Sentence> ,	150
Omission points	<Sentence> ...	800

* <Sentence> = [a-zA-Z\']+

4.3 Body

The character's body is modelled and animated using a commercial animation package. The animations are imported frame-by-frame and converted to the internal geometry format.

Each animation represents a specific gesture. There are three kinds of possible gestures: gestures that result from the internal emotional state, gestures explicitly indicated in the story (using *behaviour* tags) and gestures implicitly used during speech.

4.3.1 Emotional Gestures

At the current developmental state, the emotional gestures affect only the character's face. For demonstration purposes we defined only two facial animations (happy and sad) that are related with the *happiness* threshold.

Figure 4 - The character is happy

Figure 5 - The character is sad

4.3.2 Explicit Gestures

A set of iconic gestures is pre-defined in the animation database: *big*, *small*, *tall* and *short*.

Figure 6 - The character indicates something is big

Figure 7 - The character indicates something is small

Figure 8 - The character indicates someone is tall

Figure 9 - The character indicates someone is short

The writer should be careful in using *behaviour* tags to perform explicit gestures, as they only benefit the story if the performed action is coherent with the current story context.

Other character gestures (such as deictic gestures to refer to elements in the scene) will be considered in the future.

4.3.3 Implicit Gestures

To hide the absence of lip synchronization we use a single gesture that sequentially opens and closes the Storyteller's mouth. Since the speech is combined with the other gestures (specially the iconic gestures), the problem of not having lip synchronization is a little bit reduced.

5. ENVIRONMENT

The character's surrounding environment includes the set and the illumination pattern.

5.1 Set

For demonstration purposes, two sets were chosen to immerse the character. These two sets are briefly described in Table 5.

Table 5 – The chosen sets

Set	Description
House	This set represents the inside of a normal house, with some furniture and a fireplace. Decorative paintings exist in the wall, as well as a window showing the outside.
Street	This set represents a city street. An old building, an alley with a garbage container, two streetlights and a garage.

The sets can be toggled at any time, during the story, by using the *set* tags. New sets could be added to the application, creating a series of possible environments in which to immerse the character.

5.2 Illumination

We created two illumination patterns that can be applied to both sets, although the lights that define these two patterns may differ from one set to the other. Table 6 briefly describes the implemented illumination patterns.

Table 6 – The illumination patterns

Illumination pattern	Description
Day	Abundant and natural light. Blue sky can be seen. This is the default illumination pattern.
Night	Less light. Sunset sky can be seen. This illumination pattern includes the use of spotlights.

Again, new illumination patterns could be created to add more expressiveness to the application. For example, a new pattern entitled "Creepy" could be defined, allowing the writer to use it as a mean to stress out a particularly scary part of the story.

5.3 Set / Illumination Pattern Swap

To achieve smooth transitions between different sets and illumination patterns we use a "fader" that acts like a theatre curtain. In its normal state, the fader is transparent. This causes the scene to be completely visible. When there is the need to perform a set exchange (or an illumination pattern exchange), the visualization window starts to fade out, becoming black and hiding the scene. At this time, the set and / or the illumination pattern is exchanged, and when it is ready, the visualization window fades in, becoming transparent, and allowing the user to view the scene again.

6. RESULTS

The overall goal was achieved, but the Storyteller has some limitations that should be improved:

- The TTS system used does not provide a great deal of flexibility when it comes to using its parameters to express the emotions we intend. The character's voice seems to be more synthetic than we had hoped for.
- The bodily expression is understandable, but limited by the number of available animations. The character would be more expressive if we could combine gestures in real-time.
- From time to time, the absence of lip synchronization is very noticeable.

7. CONCLUSIONS

The overall approach (architecture and module design) seems to be adequate for the intended purpose.

Note that the current state of development of the project does not introduce a great degree of autonomy. In reality, the character is explicitly and externally controlled by the input file, which works almost as a scripting language. Naturally, the character's autonomy and sensitivity to context will be further developed, as an obvious evolution for the storytelling character.

The Storyteller has also a few limitations imposed by both the TTS system and the character engine that decrease the desired expressiveness.

8. FUTURE WORK

A series of aspects could be improved and added to the storyteller:

- Increasing the autonomy of the storytelling character.
- Replacing the TTS with an affective speech system (such as the one proposed by Cahn [3] or the one provided by the SAFIRA project [1]).
- Enhancing the character engine to allow the composition of animations.

- Improving the synchronization between lip movement and the character's voice.
- Adding new sets, illumination patterns, and character gestures to enlarge the available database and to add more semantic richness.

We intend to add the Storyteller the ability of perceiving its audience and react to external stimulus (for example, perceiving the user face and its expressions, and adjust the storytelling process accordingly).

We are also planning to extend the scripting language to enable control over the narrative. In fact, we would like to control not only the storyteller and the sets, but also the virtual actors present in the story. That way, it will be possible to have and compare different experiences of storytelling.

Further, we are considering embedding the Storyteller into applications such as Teatrix [9]. To do that, both the tags and the text need to be generated automatically from within the application.

9. REFERENCES

[1] André, E., Arafa, Y., Gebhard, P., Geng, W., Kulessa, T., Martinho, C., Paiva., A., Sengers, P., and Vala, M. SAFIRA Deliverable 5.1 – Specification of Shell for Emotional Expression, 2001. http://gaiva.inesc.pt/safira

[2] Badler, N.I., Zhao, L., and Noma, T. Design of a Virtual Human Presenter.

[3] Cahn, J.E. The Generation of Affect in Synthesized Speech. Journal of the American Voice I/O Society, vol. 8, pp. 1-19, 1990.

[4] Cassell, J. Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents.

[5] Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., and Yan, H. Embodiment in Conversational Interfaces: REA.

[6] Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., and Yan, H. Human Conversation as a System Framework: Designing Embodied Conversational Agents.

[7] Churchill, E.F., Cook, L., Hodgson, P., Prevost S., and Sullivan, J.W. "May I Help You?": Designing Embodied Conversational Agent Allies.

[8] Druin, A., Montemayor, J., Hendler, J., McAlister, B., Boltman, A., Fiterman, E., Plaisant, A., Kruskal, A., Olsen, H., Revett, I., Schwenn, T., Sumida, L., and Wagner, R. Designing PETS: A Personal Electronic Teller of Stories.

[9] Paiva, A., Machado, I., and Prada, R. Heroes, Villains and Magicians: Dramatis Personae in Virtual Environments. In Intelligent User Interfaces, ACM Press, 2001.

[10] Scherer, K.R. Emotion effects on voice and speech: Paradigms and approaches to evaluation. Presentation held at ISCA Workshop on Speech and Emotion, Belfast, 2000.

10. ACKNOWLEDGEMENTS

The authors would like to thank Marco Costa and Fernando Rebelo for the artwork, and André Vieira, Filipe Dias, José Rodrigues and Bruno Araújo for their help, ideas and comments.

This work has been partially supported by the EU funded SAFIRA project number IST-1999-11683.

Ana Paiva has also been partially supported by the POSI programme (do Quadro Comunitário de Apoio III).