Telling Stories with a Synthetic Character: Understanding Inter-modalities Relations

Guilherme Raimundo, João Cabral, Celso Melo, Luís C. Oliveira, Ana Paiva
guilherme.raimundo@tagus.ist.utl.pt; ana.paiva@inesc-id.pt

INESC-ID
Avenida Prof. Cavaco Silva - Taguspark
2780-990 Porto Salvo, Portugal

Abstract. Can we create virtual storytellers that have enough expressive power to convey a story? This paper presents a study comparing the storytelling ability between a virtual and a human storyteller. In order to evaluate it, three means of communication were taken into account: voice, facial expression and gestures. One hundred and eight students from computer engineering watched a video where a storyteller narrated the traditional Portuguese story entitled "O Coelhinho Branco" (The little white rabbit). The students were divided into four groups. Each of these groups saw one video where the storyteller was portrayed either by a synthetic character or a human. The storyteller's voice, no matter the nature of the character, could also be real or synthetic. After the video display, the participants filled a questionnaire where they rated the storyteller performance. As expected the synthetic versions used in the experiment obtained lower classifications than their natural counterparts. The data suggests that the gap between synthetic and real gestures is the smallest while the synthetic voice is the furthest from its natural version. An interesting result was that the classification of the facial expression is affected by the nature of the voice.

1 Introduction

A strong bond exists between storytelling and human society. All of us have the need to express ourselves and to communicate our experiences to others. We accomplish this through storytelling. Thanks to storytelling, our ancestors' culture was passed from generation to generation giving us meaning in the world [18]. With the new technologies it is unavoidable that such way of communication is adapted to fit this new paradigm. Several researchers have been focusing in the interpretation of the audience reaction while others look into the storyteller's expression of the story [20][16]. There are still those who focus in the automatization of the process of creating a story [1][4]. Various studies exist that evaluate synthetic characters in environments such as E-retail [22] and education [2]. However, to the best of our knowledge, there is still no in depth study that attempts to compare the performance between a human and a synthetic storyteller. This study tries to give some insight into this issue by evaluating a

storytelling system in relation to a human actor.
In order to measure the storyteller performance we considered three means of communication used in the storytelling: gestures, facial expression and voice. Furthermore, we considered that story understanding, conveying of emotion, believability and satisfaction [19], are essential aspects in the performance of a storyteller. To compare performance between storytellers, four videos were created. In each of these, the storyteller character could be a human actor or a 3D character. The storyteller's voice could also be from the human actor or synthesized. Each study participant visualized one of these videos and rated each communication mean in the four mentioned aspects. From the experiment we wanted to investigate if there were significant differences between the storytelling ability of our system and the human actor. Also, we aspired to determine the influence of the character and voice nature in the result of the storytelling rating.

2 Method

Each participant visualized one of four videos where the storyteller narrated the traditional Portuguese story "O Coelhinho Branco". After the visualization, the performance of the storyteller was evaluated through a questionnaire.

2.1 Design

Two independent variables were used: Character and Voice. Each of these is composed of a real level (human actor, human actor voice) and a synthesized level (3D character, synthesized voice). The real version uses the recording of a human actor while telling the story. The synthetic version uses a 3D character that is a blend between an old man and a tweenie [21]. Both characters can be seen on Figure 1.

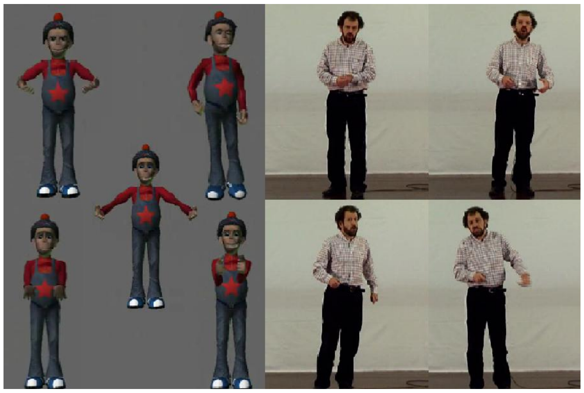

Fig. 1. Examples of the synthetic character and human actor telling the story.

So that both versions of the character supplied the same knowledge to the participants, the semantic information transmitted by human actor gestures and facial expression was annotated. This annotation was then used in the creation of the synthetic character gestures and facial expressions. Regarding facial expression, the six basic emotions of Ekman [8] (Joy, Sadness, Anger, Surprise, Disgust and Fear) were taken into account. Particular facial area movements, such as the eyebrows, that are used to convey or reinforce currently spoken information, were also annotated. Concerning gestures, the focus of annotation resided in the gesticulation, i.e., in the unconscious and idiosyncratic movement that carries some communicative meaning [11]. To bring this annotations to life a character animation engine was created. This engine pays special attention to the processes of animation of facial and body expression in humanoid characters. In fact, it is not only capable of playing hand made deterministic animation but also allows a finer and more expressive control of isolated character parts.

Facial Expression - Here, the fine animation control is obtained through the use of control parameters. A control parameter is a value that has a maximum, minimum and associated semantic information. This information dictates what happens when the value of the parameter varies. For example the intensity of contraction of the major zygomatic muscle, the rotation angle of the left eye or the degree of joy the face expresses. The existing parameters can be divided into two sets: atomic and group. The parameters that are atomic contain all needed information to create the desired deformation. The engine allows the use of three types of atomic parameters: pseudo-muscular, transformation and skinning. The pseudo-muscular parameter follow a deformation model based in Waters model [23] and emulate the behavior of the contraction of a muscle under the skin. In the used storyteller character 37 pseudo-muscles were used. The transformation parameters simply apply a geometric transformation to a given geometric object. This type of parameters are used to control the rotation of the synthetic character eyes. The skinning parameters use a known animation technic that uses weighted mesh connected to virtual bones. This type of parameter is used for the tongue and jaw movement of the synthetic character. The group parameters (non-atomic) are used, as the name mentions, to group several control parameters together. These are usually used to create abstractions of resulting deformations from several parameters. Emotional expressions and visemes are two examples of group parameters. Visemes are the facial displays when a given phoneme is spoken. An interpolation between consecutive visemes is executed for viseme co-articulation.

Gestures - Gestures in the synthetic character are based in a articulated modeled that is structured in a hierarchic architecture with three layers (similar to [3] and [17]): geometry, animation and behavior. The model supports deterministic animation based in keyframes and non-deterministic animation that is dynamically generated in real time through the use of inverse kinematics. Considering the deterministic animation, the geometry layer defines a skeleton, inspired in the human, that is composed of 54 bones; the animation layer allows the execution and combination of animations which are defined over subsets of the subjacent skeleton and keyframed based.; the behavior layer supplies scripting abilities which allow the execution of complex animation sequences. Regarding the non-deterministic animation, the geometric layer makes use of robotic manipulator members with 6 rotation junctions; the animation layer implements the primitives of direct kinematic, inverse kinematic and inverse velocity; in the behavioral layer the scripting is extended in order to allow the new primitives of the non-deterministic animation.

The gestures model permits gesticulation animation, i.e., the type of unconscious idiosyncratic movement with communicative meaning that occurs in the context of a dialogue or narration [11]. The model is restricted to the upper body since, according to McNeill [11], gesticulation occurs predominantly through the arms and hands. Concretely, the model, is built upon the deterministic and non-deterministic animation architecture allowing real time gesticulation defined as an arbitrary sequence of positions, orientations and shapes of the hands.

For hand shapes the model allows the use of most static shapes from the Gestural Portuguese Language [9]. Regarding hand's orientation and position, the model allows, through the use of inverse kinematic, the animation of arbitrary trajectories in the space that surrounds the synthetic character.

The gesture expression of the synthetic character in the story corresponds to the application of a recording algorithm [10] for gesticulation transcription to the human actor video and to the later conversion of this annotation into animation scripts. When the gesticulation done by the human actor was too complex, keyframed animations were created.

Voice - In order to command the facial movements of the synthetic character in synchrony with the speech of the human actor, the natural phonetic signal was annotated. This process was performed in a semi-automatic manner. Since the actor wasn't obliged to follow a strict script, after the recording of the video, the performed story was transcribed. From this transcription several levels of automatic analyses were made that allowed to determine a possible phonetic sequence for the text. Following this process, also in a automatic manner, the sequence was aligned with the original speech signal [**?**]. Then it was considered the possibility that the speaker produced alternative pronunciations to the ones determined by the text analyses [14][15] resulting in a more accurate estimative of the performed phonetic sequence. Finally, the outcome of the automatic analyses was manually verified and some boundaries of phonetic segments were corrected.

For the synthesis of the synthetic voice it was also necessary to guarantee the synchronism between the speech signal and the video sequence. In order to achieve this goal it was necessary to impose that the duration of the synthetic phonetic segments was equal to the originals. Since the determination of the contour of the fundamental frequency is intimately related with the rhythm attribution,

it was opted to impose the actor produced contour to the synthetic voice. The synthetic voice creation is made with a diphone synthesizer based in Linear Predictive Coding with a male voice. The synthesizer was developed at INESC-ID and uses as reference the original durations and produced speech with constant fundamental frequency.
We have at our disposal speech synthesizers with selection of variable dimension units which supply better quality synthesis. However, they were not used because they do not allow the same flexibility for the production of the synthetic signal. This signal was processed later on in order to have the same intonation of the speech produced by the human actor. Since the actor used a falsetto voice and the synthesizer uses a neutral voice, the variation of the fundamental frequency necessary to be used in the synthetic voice surpassed many times the 1.5 factor which usually is considered as an acceptable limit of distortion. In order to minimize this effect, a new technic named PSTS, was developed to alter the duration and fundamental frequency of the speech signal [5][6]. This technic also allows changing the speech signal parameters associated with the vibration form of the glottis. This is important for the production of speech with certain emotions. The way this parameters are changed in order to transmit those emotions is a current working topic [7]. Therefore, in this present study, the emotions present in the speech signal are transmitted solely by the variation of the rhythm and intonation.

The experiment followed an independent sample design, with each participant being assigned to a unique combination of the independent variables. There are 12 dependent variables in the experiment corresponding to the combinations of the 3 communication means (gestures, facial expression and voice) with the 4 analyzed aspects (story understanding, conveying of emotion, believability and satisfaction). These were measured through the use of a questionnaire explained bellow.

2.2 Participants

The study had the participation of 108 students of computer engineering from Instituto Superior Tecnico. From them, 89 were male and 19 female. Their ages varied between 18 and 28 years old with an average of 21 years and 10 months. The participants had no previous knowledge of the experiment objectives, knowing only that it was related to storytelling in virtual environments.

2.3 Material

For the video visualization, computers with 19" LCDs were used along with headphones for the audio. Each video had the duration of 7 minutes and 29 seconds and showed one level of each independent variable. For the evaluation of the video by the participants a questionnaire was created. This questionnaire is composed of 12 statements that result from the combination between the communication means (gestures, facial expression and voice) and considered aspects (story understanding, conveying of emotion , believability and satisfaction). Therefore, each statement is an assertion about one aspect of one of the means of communication. The participants rated the statements through a Likert scale with values between 1 and 7. Choosing the value 1 meant total disagreement with the statement, value 4 neither disagreement nor agreement with the statement and value 7 total agreement with the statement. Although the order of the statements in the questionnaire was obtained in a randomly fashion, we show them here sorted by communication mean and considered aspects.

1. The facial expressions helped in the understanding of the story
2. The storyteller's face expressed the story emotions
3. The facial expressions were believable
4. I liked the facial expressions
5. The gestures helped in the understanding of the story
6. The gestures expressed the story emotions
7. The gestures were believable
8. I liked the gestures
9. I understood everything the storyteller said
10. The voice expressed adequate emotions regarding the story
11. The voice was believable
12. I liked the voice

2.4 Procedure

The four possible combinations between the independent variables formed the sample groups displayed in Table 1.

	Real Character	Synthetic Character
Real Voice	RSRV	SSRV
Synthetic Voice	RSSV	SSSV

Table 1. Sample Groups

Each participant was assigned randomly to one of the four groups complying only with the restriction of equal participant numbers between groups. Thus, each sample group was constituted of 27 elements. At the beginning of each visualization the questionnaire was briefly explained to the participant. It was mentioned that the participant should read the questionnaire introduction before the video visualization and that he should fill out the rest of the questionnaire after the visualization.

3 Results

This section presents the results obtained by the carried out study. The data is depicted in tables 2, 3 and 4. Table 2 shows the percentage of negative (disagreement with the statement), neutral (neither agreement nor disagreement with the statement) and positive (agreement with the statement) ratings. This table displays the data organized by independent variable and video. Table 3 reveals the results obtained through an analysis of variance. The statistical test used is a two-way non-parametric analysis of variance described in [12]. The test is similar to a Kruskal-Wallis test extending it to consider two independent variables and possible interaction between them. Finally, Table 4 shows the results of a Mann-Whitney test between the RSRV and SSRV groups. With this grouping we only consider the variation of the nature of the character. This last test was created to isolate the effect of the synthesized voice over the rating of the facial expression. In all tests it was considered that a difference was significant for $p < 0.05$.

Since the amount of gathered data is relatively high we opted to present the main results in a hierarchical order. First we will consider the variation of the independent variables. For each level of the independent variables we will then focus on a particular communication mean. Within each communication mean we present the results for each considered aspect. Second, we take into account the interaction effect between the independent variables. Last we present the results from the Mann-Whitney analysis.

Differences between Real and Synthetic Character

Facial Expression - Significant differences were found in the rating of facial expression for story understanding ($H = 7.48$, $df = 1$, $p = 0.006$), conveying of emotion ($H = 7.13$, $df = 1$, $p = 0.008$), believability ($H = 12.79$, $df = 1$, $p < 0.001$) and satisfaction ($H = 10.46$, $df = 1$, $p = 0.001$). In all aspects the synthetic character facial expression received lower ratings than the human actor.

Gestures - In the rating of the gestures a significant difference was found for the believability aspect ($H = 8.26$, $df = 1$, $p = 0.004$), having the synthetic character less believable gestures than its real counterpart. No significant differences were found for story understanding ($H = 1.12$, $df = 1$, $p = 0.290$), conveying of emotion ($H = 1.61$, $df = 1$, $p = 0.204$) and satisfaction ($H = 3.66$, $df = 1$, $p = 0.056$) aspects.

Voice - As expected, no significant differences were found for story understanding ($H = 2.40$, $df = 1$, $p = 0.121$), conveying of emotion ($H = 0.001$, $df = 1$, $p = 0.979$), believability ($H = 0.105$, $df = 1$, $p = 0.746$) and satisfaction ($H = 0.004$, $df = 1$, $p = 0.950$) of the voice when varying the nature of the character.

Differences between the Real and Synthetic Voice

Facial Expression - Significant differences were found in the rating of facial expression for story understanding ($H = 3.89$, $df = 1$, $p = 0.049$), conveying of emotion ($H = 6.64$, $df = 1$, $p = 0.010$), believability ($H = 5.87$, $df = 1$, $p = 0.015$) and satisfaction ($H = 9.92$, $df = 1$, $p =0.002$). In all aspects, facial expression received lower ratings when the synthesized voiced was used.

Statement#	%	Character		Voice		Video			
		Real	**Virtual**	**Real**	**Virtual**	**RSRV**	**SSRV**	**RSSV**	**SSSV**
1	Positive	70.4	42.6	63.0	50.0	77.8	48.1	63.0	37.0
	Neutral	13.0	14.8	13.0	14.8	14.8	11.1	11.1	18.6
	Negative	16.6	42.3	24.0	35.2	7.4	40.8	25.9	44.4
2	Positive	83.3	59.2	77.8	64.8	85.2	70.4	81.5	48.2
	Neutral	11.1	13.0	9.2	14.8	7.4	11.1	14.8	14.8
	Negative	5.6	27.8	13.0	20.4	7.4	18.5	3.7	37.0
3	Positive	77.8	50.0	72.2	55.6	81.5	63.0	74.1	37.0
	Neutral	16.6	13.0	14.8	14.8	18.5	11.1	14.8	14.8
	Negative	5.6	37.0	13.0	29.6	0.0	25.9	11.1	48.2
4	Positive	75.9	44.4	70.4	50.0	81.5	59.3	70.4	29.6
	Neutral	11.1	24.1	16.6	18.5	11.1	22.2	11.1	26.0
	Negative	13.0	31.5	13.0	31.5	7.4	18.5	18.5	44.4
5	Positive	83.3	75.9	79.6	79.6	85.2	74.1	81.5	77.8
	Neutral	7.4	5.6	5.6	7.4	7.4	3.7	7.4	7.4
	Negative	9.3	18.5	14.8	13.0	7.4	22.2	11.1	14.8
6	Positive	87.0	64.8	72.2	79.6	81.5	63.0	92.6	66.7
	Neutral	7.4	11.1	9.3	9.3	11.1	7.4	3.7	14.8
	Negative	5.6	24.1	18.5	11.1	7.4	29.6	9.7	18.5
7	Positive	75.9	57.4	66.7	66.7	74.1	59.3	77.8	55.6
	Neutral	13.0	14.8	9.2	18.5	14.8	3.7	11.1	25.9
	Negative	11.1	27.8	24.1	14.8	11.1	37.0	11.1	18.5
8	Positive	77.8	63.0	63.0	77.8	66.7	59.3	88.9	66.7
	Neutral	13.0	16.6	16.6	13.0	22.2	11.1	3.7	22.2
	Negative	9.2	20.4	20.4	9.2	11.1	29.6	7.4	11.1
9	Positive	87.0	77.8	94.4	70.4	96.3	92.6	77.8	63.0
	Neutral	3.7	1.8	1.9	3.7	3.7	0.0	3.7	3.7
	Negative	9.3	20.4	3.7	25.9	0.0	7.4	18.5	33.3
10	Positive	81.5	83.3	94.4	70.4	92.6	96.3	70.4	70.4
	Neutral	7.4	9.3	3.7	13.0	7.4	0.0	7.4	18.5
	Negative	11.1	7.4	1.9	16.7	0.0	3.7	22.2	11.1
11	Positive	68.5	74.1	88.9	53.7	92.6	85.2	44.5	63.0
	Neutral	11.1	9.2	11.1	9.3	7.4	14.8	14.8	3.7
	Negative	20.4	16.7	0.0	37.0	0.0	0.0	40.7	33.3
12	Positive	53.7	51.9	77.8	27.8	81.5	74.1	25.9	29.6
	Neutral	16.7	14.8	18.5	13.0	18.5	18.5	14.8	11.1
	Negative	29.6	33.3	3.7	59.2	0.0	7.4	59.3	59.3

Table 2. Percentage of Positive, Neutral and Negative ratings

Gestures - As shown in Table 3, there are no significant differences in the evaluation of gestures when varying the nature of the voice.

Voice - There is a high significant difference ($p < 0.001$) for all considered aspects of the voice with the synthesized voice having lower ratings than the real voice.

Interaction Effect between Character and Voice

As can be seen in Table 3 there is no significant interaction effect between Character and Voice for all statements. Though, it should be noticed that statement 9, concerning the story understanding trough the voice communication mean, has an interaction effect value ($p = 0.059$) close to significant.

Stat.#	**Source**	*p*	**Mean Rank**		**Stat.#**	*p*	**Mean Rank**	
			Real	**Synthetic**			**Real**	**Synthetic**
1	Character	0.006	62.56	46.44	7	0.004	62.90	46.10
	Voice	0.049	60.31	48.69		0.142	50.21	58.79
	Character*Voice	0.557				0.675		
2	Character	0.008	62.31	46.69	8	0.056	60.02	48.98
	Voice	0.010	62.04	46.96		0.082	49.49	59.51
	Character*Voice	0.453				0.693		
3	Character	<0.001	65.01	43.99	9	0.121	48.79	50.21
	Voice	0.015	61.62	47.380		<0.001	65.60	43.40
	Character*Voice	0.512				0.059		
4	Character	0.001	64.03	44.97	10	0.979	54.57	54.43
	Voice	0.002	63.78	45.22		<0.001	66.33	42.67
	Character*Voice	0.640				0.928		
5	Character	0.290	57.47	51.53	11	0.746	53.55	55.45
	Voice	0.919	54.21	54.79		<0.001	68.00	41.00
	Character*Voice	0.510				0.887		
6	Character	0.204	58.17	50.83	12	0.950	54.32	54.69
	Voice	0.626	53.09	55.91		<0.001	71.32	37.69
	Character*Voice	0.612				0.866		

Table 3. Two-Way Non-Parametric ANOVA Test Results

		Statement #1	**Statement #2**	**Statement #3**	**Statement #4**
Mean Rank	RSRV	32.15	29.80	31.67	31.44
	SSRV	22.85	25.20	23.33	23.56
Mann-Whitney U		239	303	252	258
p (2-tailed)		0.026	0.270	0.045	0.058

Table 4. Mann-Whitney Test Results

Difference between Real and Synthetic Character only considering the Real Voice sample groups

Facial Expression - There is a significant difference for the story understanding ($U = 239$, $p = 0.026$) and believability ($U = 252$, $p = 0.045$) aspects, with the character obtaining lower ratings for the 3D character. No significant difference was found for conveying of emotion ($U = 303$, $p = 0.270$) and satisfaction ($U = 258$, $p = 0.058$) aspects.

4 Discussion

4.1 Analysis

In a general manner it can be concluded that the synthetic versions used in the experiment obtain worse classifications than their real counterparts. The data suggests that the synthesized gestures are the closer to the human version and that the synthesized voice has the furthest distance to the performance of the human actor. An interesting result is that the rating of the facial expression is affected not only by its real or synthetic nature but also by the nature of the voice used.

Facial Expression - For all dependent variables, the synthesized facial expression has a significant lower rating than the real one. Of particular interest is that the rating of this communication mean is strongly affected not only by the visual expression but also by the voice. In fact, the use of synthesized voice has a significant negative effect when rating the facial expression. To isolate this effect, a statistical test was performed where only the human actor voice was used. With it we concluded that for the expression of emotions and for the satisfaction of this communication mean, the difference between the real and synthetic storyteller was no longer significant. This fact suggests that these rank averages in particular are more affected by the synthetic voice. As is shown in statement 2 of Table 2, by only considering the human voice we have that the positive percentage rating is of 85.2% for the RSRV video and of 70.4% for the SSRV video. This 14.8% difference is a bit less than the half of the difference between the videos RSSV and SSSV where the percentage of positive ratings drops from 81.5% to 48.2%. By consulting statement 4 of the same table we encounter a similar behavior in what concerns the rating of facial expression satisfaction.

Gestures - Regarding gestures, only one significant difference was found in the rating of its believability. In this case the synthetic storyteller presents worse performance than the human actor. In the remaining ratings the data suggests that the synthetic gestures have a close performance to the real ones. It is also worthy of notice that gestures rating have always a majority of positive ratings (statements 5, 6, 7 and 8 of Table 2). Similarly to what happens in the facial expression, gestures also seem to be affected by the nature of the used voice, but this time in an inverse manner. Positive gestures ratings percentages have an increase or stay on the same value when the synthesized voice is taken into account. Unfortunately we did not achieve a significant difference when varying the voice nature, being the satisfaction rating the closest one to achieve such difference with a $p = 0.082$.

Voice - Through an analysis of the results, on the statement "I understood everything the storyteller said", we notice that there is a very close to significant value ($p = 0.059$) for the interaction between independent variables. This lead us to believe that this statement is not measuring what we intended. Therefore, we discarded this statement for analysis. In the remaining aspects, the voice was the medium that had a clearer significant difference between the real and the synthetic versions ($p < 0.001$), having the real voice higher ratings than its counterpart. Nevertheless, only the satisfaction regarding the synthetic voice obtained a majority of negative ratings (see statement 12 of Table 2). Both the emotion and believability aspect of the synthetic voice gathered a majority of positive ratings with values of 70.4% and 53.7% respectively.

4.2 Study Limitations

Like all studies this one is not without its limitation. By using subjective classifications from participants we may not be achieving the desired precision or obtaining data that is not representative of what we want to measure. Another problem at hand is that the sample used is only representative of the computer engineering students from Instituto Superior Tecnico population. To solve this issue the study should be extended to consider a larger number of participants with higher diversity.

4.3 Future Work

As future work we propose the creation of a story where the information conveyed by each communication mean is clearly defined and controlled. It would be then possible, through the use of exclusive information, to create a better questionnaire to assess what knowledge of the story the listener obtained from each of the communication means. The same approach could be used to measure the performance of emotive information transmission. Indirect methods, such as the comparison with real situations, could be used to measure believability. However, the creation of a precise measure of believability is a complex task that gives origin to a debate on its own. We consider that the impact of the voice in the facial expression appreciation is something worthy of future studies. A dedicate study may conclude if the relation uncovered by the study really exists and if it has the direction that the data indicated. Another topic for a future study is related to the acceptance of the synthetic voice. Contrarily to what happens with the figurative representation of the character, it appears that the users are expecting that the synthetic character to have a human voice. This experience is certainly due the long exposure of animated character which speech is borrowed from human performers. This means that we accept figurative representations to stylize humans, animals and even objects but we demand that when they speak they produce a human-like speech. The fact that the figurative representation does not possess the physical mechanism that allow the production of such acoustic signal, does not seem to affect in its believability. A synchronized labial movement with the speech suffices for the viewer to expect a human like vocal signal. A study concerning the acceptable level of distortion or of a stylized synthetic speech that is accepted by the viewer are certainly good topics for further research.

5 Acknowledgments

This project was partially funded by the Portuguese Foundation for Science and Technology (POSI/SRI/41071/2001) and the EU funded Humaine Network of Excellence contract number 507422 of the IST programme.

References

1. Aylett, R.; Louchart, S.; Dias, J; Paiva, A.; Vala, M.: FearNot! - An Experiment in Emergent Narrative IVA 2005, 305-316

2. Baylor, A.; Kim, Y: "Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role." International Conference on Intelligent Tutoring Systems, Macei, Brazil. (2004)
3. Blumberg, B.; Galyean, T.: Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments Computer Graphics (SIGGRAPH 95 Proceedings), 30(3):47-54, 1995
4. Brooks, K: Do Story agents use rocking chairs ICM, 1997
5. Cabral, J.; Oliveira, L.: Pitch-Synchronous Time-Scaling for High-Frequency Excitation Regeneration, Interspeech 2005, Sep. 2005
6. Cabral, J.; Oliveira, L.: Pitch-Synchronous Time-Scaling for Prosodic and Voice Quality Transformations, Interspeech 2005, Sep. 2005
7. Cabral, J.: Transforming Prosody and Voice Quality to Generate Emotions in Speech, MSc Thesis, IST, UTL, Jan. 2006
8. Ekman, P.: Facial Expressions in Dalgleish, T., & Power, M., Handbook of Cognition and Emotion. New York: John Wiley & Sons Ltd (1999)
9. Secretariado Nacional para a Reabilitaão e Integraão das Pessoas com Deficincia: Gestuário Lngua Gestual Portuguesa, 5 edi co
10. Melo, C.; Paiva A.: "A Story about Gesticulation Expression" (under submission)
11. McNeill, D.: Hand and Mind: What gestures reveal about thought, The University of Chicago Press, 1992
12. Maroco, J.; Bispo, R.: "Estatstica aplicada s cincias sociais e humanas", Climepsi Editores, 2003, pag. 249-253, ISBN-9727960650
13. Paulo, S.; Oliveira, L.: Improving the Accuracy of the Speech Synthesis Based Phonetic Alignment Using Multiple Acoustic Features, Computational Processing of the Portuguese Language - Proc. of the 6th Intl. Workshop, PROPOR 2003, Jun. 2003 , pp. 31-39 , Springer-Verlag, Heidelberg
14. Paulo, S.; Oliveira, L.: Generation of Word Alternative Pronunciations Using Weighted Finite State Transducers, Interspeech 2005, Sep. 2005 .
15. Paulo, S.; Oliveira, L.: Reducing the Corpus-based TTS Signal Degradation Due to Speaker's Word Pronunciations, Interspeech 2005, Sep. 2005 .
16. Pelachaud C.; Maya, V.; Lamolle, M.: Representation of Expressivity for Embodied Conversational Agents Workshop Balanced Perception and Action, Third International Joint Conference on Autonomous Agents & Multi-Agent Systems, New-York, julho 2004
17. Perlin, K.; Goldberg, A.: Improv: A system for scripting interactive actors in virtual worlds Computer Graphics (SIGGRAPH96), 30:205-218, 1996
18. Roemer, M: "Telling Stories : Postmodernism and the Invalidation of Traditional Narrative", Rowman & Littlefield Publishers, Inc., 1995, ISBN-0847680428
19. Sawyer, R.: "The way of the Storyteller", Penguin Books, 1942
20. Silva A.; Raimundo G.; Paiva A., de Melo C.: To tell or not to tell...Building an interactive virtual storyteller, in Proceedings of the Language, Speech and Gesture for Expressive Characters Symposium, AISB 2004, Leeds, Reino Unido, Abril de 2004
21. Online reference - http://www.bbc.co.uk/cbeebies/tweenies/ - Last access 23/02/2006
22. McBreen H.; Shade P; Jack M; Wyard P.: "Experimental assessment of the effectiveness of synthetic personae for multi-modal e-retail applications." Agents 2000: 39-45
23. Waters, K.: "A muscle model for animation three-dimensional facial expression." In Proceedings of the 14th Annual Conference on Computer Graphics and interactive Techniques M. C. Stone, Ed. SIGGRAPH 1987. ACM Press, New York, NY, 17-24.