Metrics for Character Believability in Interactive Narrative

Paulo Gomes[1], Ana Paiva[2], Carlos Martinho[2], and Arnav Jhala[1]

[1] UC Santa Cruz, Santa Cruz,
CA, 95064, USA
[2] Instituto Superior Técnico, Technical University of Lisbon,
2744-016 Porto Salvo, Portugal

Abstract. The concept of character believability is often used in interactive narrative research hypothesis. In this paper we define believability metrics using perceived believability dimensions and discuss how they can be accessed. The proposed dimensions are: behavior coherence, change with experience, awareness, behavior understandability, personality, visual impact, predictability, social and emotional expressiveness.

Keywords: believability, metrics, emotion.

1 Introduction

In the beginning of the 19th century, the english poet Samuel Taylor Coleridge coined the term *suspension of disbelief* [3]. The term referred to the mental state in which the reader of a poetic piece could regard a supernatural, or simply romantic, character as real, regardless of characteristics out of the ordinary. When explaining his motivation for *Lyrical Ballads* [4], Coleridge expresses his desire to write in "semblance of truth" and to spur the reader's imagination, clouding what would, at first glance, and without context, be regarded as unrealistic.

Since Coelridge's *Biographia Literaria*, the term has evolved: the concept of character was generalized to a fictional situation; and the term now encompasses any art form, and not specifically poetry. One such art form is Animation, that in the 1930's saw great technical, and aesthetic progress, in the hands of Walt Disney Studio artists. Thomas and Johnston would describe the animation principles learned by these artists in the seminal *The Illusion of Life: Disney Animation* [19]. In this book, they explain how animated characters can give the illusion of being alive, of having motivations, of thinking and acting accordingly. Later on, Lasseter identified the possibility of applying the lessons of traditional animation to this new area [9].

In the nineties, computer science researchers in the field of autonomous agents began to analyze how the artistic principles of animated characters could be used to design believable agents that gave "the illusion of life". Researchers from Carnagie Mellon who were part of the OZ project made a significant thrust in this direction [2][11][13]. Loyall dissected the definition of believable agent and proposed several requirements, many of them related to models of personality [11].

H. Koenitz et al. (Eds.): ICIDS 2013, LNCS 8230, pp. 223–228, 2013.

By personality he considered "all of the particular details - especially details of behavior, thought and emotion - that together define the individual" (p. 16). Ortony [16] proposed a definition for believable agents more centered around emotion. He considered that there should be consistency in the way agents evaluate events, and how this evaluation influences their emotional state. Agent believability was also analyzed in more specific contexts, such as pedagogical agents by Lester and Stone [10]. They defined believability as the identification, from the user or spectator, of an agent's goals, beliefs and personality.

These definitions of believability provide guidelines to AI designers on how to design systems that support believable characters. For instance, the interactive drama Façade [14] combines a structured narrative with simulated autonomous agents, and uses such guidelines to shape character behavior. Specific implementations of AI systems focus on different aspects of believability. This makes it difficult to evaluate the degree to which characters developed in different AI architectures and behavior models are believable.

This paper presents initial work in defining metrics for evaluating believable virtual characters in interactive narratives. Specifically, we are interested in analyzing how an audience unfamiliar with the abstract concept of believability can provide feedback on the believability of characters whose behavior is defined by a computational system. The idea is to tap into their interpretation of the narrative discourse generated. The ultimate goal is to create metrics to help researchers assess how a computational system is contributing to the final believability of characters in interactive narratives.

There have been efforts to measure different notions of believability: in relation to expectations [12], in the context of story generation [18], connected with trustworthiness [1], for opponents in a first person shooter [8], through empathy [17]. Nonetheless, these works to not evaluate a broad notion of believability considering consistency, emotions, personality, and other elements deemed important to believability [11][16][10]. .

In this article we will identify dimensions of believability an audience can report about, describe how to use them to access overall beliveability, discuss limitations of the methodology, and propose future work.

2 Dimensions of Believability

The ultimate goal of our metrics is to measure perceived character believability. However, directly asking an audience how believable a character is can be an ambiguous question. Unless, the audience is familiar with the notion of illusion of life, the answer will probably not reflect this concept. Hence, we propose a measure that uses several believability dimensions contributing to the overall perception of believability. In this way, participants are asked about more objective aspects of the agent. Here are the believability dimensions that we consider an audience can self-report about:

- **behavior coherence** : according to Ortony [16] coherence is a crucial element of believability. An audience will see the character's behavior, and not explicitly its internal state, so they can be asked about the coherence of the first.
- **change with experience** : the agent's change is an element referenced in [11]. In the context of interactive narrative, it can be related to Mckee's idea of story event, a significant change in a life value of a character [15]. These events are essential building blocks of a classical plot arch.
- **awareness** : agents should show they perceive the world around them. This dimension can be mapped to Lester and Stone's situated liveliness [10] as well as Loyall's reactive and responsive elements [11].
- **behavior understandability** : in Ortony's definition of believability [16], it is implicit that participants must be able to create a model of an agent's behavior motivations. Furthermore, Bates [2] points out that an agent's actions must express what it is thinking about and its emotional state. For situations in which the thought process is not explicitly shown this last sentence can be translated to: the agent's actions must express what the participant thinks the character is thinking about. But for this to happen, the spectator must be able to create a model of the character's thought process. Hence, the participant must understand the character's behavior.
- **personality** : the notion of personality is present in almost all believability definitions presented. Following Loyall's definition [11], participants should be able to identify the agent's behavior details that define it as an individual, that make it unique.
- **emotional expressiveness** : the extent to which the character expresses its emotions. The concept of emotion is mentioned by Loyall [11] and Ortony [16].
- **social** : participants should be able to identify social relationships between characters [11].
- **visual impact** : is the amount by which an agent draws our attention, and was proposed by Lester and Stone as an enhancer of believability [10].
- **predictability** : Lester and Stone also point out the importance of behavior patterns not being recognizable, specially in the context of long term interactions. Moreover, when considering variability, Ortony [16] warns for the harmful effect predictability can have on believability. However, Ortony also stated that complete lack of predictability may affect behavior coherence, and consequently believability. Thus, both extreme predictability and extreme unpredictability harm believability.

By having believability measured through factors, measurements can be more informative as design feedback. For instance, if a graphic agent scores low on awareness, authors may need to create more looking/staring animations, or provide better programmatic hooks for those animations.

3 Quantifying Believability

We want to describe how to assess an audience's perception of a character's believability through dimensions. Since believability is an internal construct to each individual we believe that the most adequate strategy would be to ask the audience to self report about their perception. Likert scales are commonly used in assessing individual subjective perceptions, thus we propose to use them by having one scale per dimension (except for emotional expressiveness that we will tackle separately). We now present templates for phrases that a participant would have to rate in which the $< X >$ field would be replaced by the name of the character being analyzed. The range boundary values would be labeled as "totally agreeing" and "totally disagreeing" with the statement. Here are the templates:

- awareness: $< X >$ perceives the world around him/her.
- behavior understandability: It is easy to understand what $< X >$ is thinking about.
- personality: $< X >$ has a personality.
- visual impact: $< X >$'s behavior draws my attention.
- predictability: $< X >$'s behavior is predictable.
- behavior coherence: $< X >$'s behavior is coherent.
- change with experience: $< X >$'s behavior changes according to experience.
- social: $< X >$ interacts socially with other characters.

For emotional expressiveness we would ask participants what emotions they believe the character was mainly expressing at specific situations. That could be assessed by having a multiple choice in which each option corresponded to a basic emotion such as anger or fear [5]. In this case a higher value would correspond to a higher frequency of correctly identified emotions. By correct we mean according to what the system was trying to express. These scales were used in previous work [7] with anecdotal evidence gathered that users understood the questions.

It is our belief that enhanced perception of the believability dimensions corresponds to a greater sense of believability, with the exception of predictability. The hypothesis that a character controlling system B promotes a higher sense of believability than a system A would be supported if all the following conditions are verified:

- No dimension (excluding predictability) is significantly higher (higher agreement) in A than B.
- System B does not have predictability values significantly closer to one of the rating extremes (totally agree or totally disagree) than A.
- The overall accuracy of character emotion identification is higher than chance $(100\%/(numberofexpressions))$.
- At least one dimension (excluding predictability) is higher in version B than version A, or version A has a predictability value significantly closer to one of the extremes than version B.

4 Conclusion

In this paper we have argued for systematic believability metrics that support research hypotheses related to interactive narrative systems. We defined the following believability dimensions: behavior coherence, change with experience, awareness, behavior understandability, personality, visual impact, predictability, social and emotional expressiveness. We described how to measure believability using phrase templates related to these dimensions. Finally, we presented a method for using these measures to support a belivability hypothesis.

A possible critique to our metrics is that the suggested dimensions have a significant degree of ambiguity to a rating audience. Nonetheless, it should still be less ambiguous than simply asking an audience to rate characters' believability. We believe that this work would serve as a strong basis for refining the measurements of believability for direct and systematic comparison of narrative systems.

As future work, we propose to validate the presented metrics. We would present participants in an experiment with non interactive animated clips of cartoons that are considered references in believability portrayal (e.g. old Disney animations, Pixar shorts). Then, ask them to rate these cartoons according to the above mentioned dimensions. Afterwards, verify if the ratings are according to what was expected, that is, dimensions were rated as high in general apart from predictability that should have medium scale values.

In our work we have focused on a sense of the term believability connected with *observed behaviors* and less on internal representation of agents' mental model. Nonetheless, the term is used with other senses. One sense is in terms of creating a believable *mental model* of agents that is coherent and consistent with the rules of the virtual environment with respect to the agent's beliefs, desires, and intentions. Another sense is in terms of *visual believability or realism*, especially regarding graphical representation and facial expressions, and in areas like robotics with physical believability of robotics rigs. Third is in terms of *believable interactions* that take in the context of communication protocols and languages (including verbal and non-verbal forms of communication) of agents make the overall interaction with human participants more natural. We suggest that researchers clearly specify what definition of believability they are using since each one entails different research goals.

Believability is an inspiring concept that has been around in the computer science context for a while. It is time to reclaim it as a measurable factor.

Acknowledgments. This work has been supported, in part, by the FP7 ICT project SIREN (project no: 258453). We would like to thank Michael Mateas for his comments on believability metrics as a design feedback tool.

References

1. Bartneck, C.: How convincing is mr. data's smile: Affective expressions of machines. User Modeling and User-Adapted Interaction 11(4), 279–295 (2001)
2. Bates, J.: The role of emotion in believable agents. Commun. ACM 37(7), 122–125 (1994), 176803
3. Coleridge, S.T.: Biographia Literaria: Biographical Sketches of my Literary Life & Opinions. Princeton University Press (1985)
4. Coleridge, S.T.: Lyrical Ballads. Penguin Classics (2007)
5. Ekman, P.: An argument for basic emotions. Cognition & Emotion 6(3-4), 169–200 (1992)
6. Frasca, G.: Video games of the oppressed (2006)
7. Gomes, P.F., Martinho, C., Paiva, A.: I've been here before! location and appraisal in memory retrieval. In: Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems, Taipei, Taiwan, pp. 1039–1046. IFAAMAS (2011)
8. Gorman, B., Thurau, C., Bauckhage, C., Humphrys, M.: Believability testing and bayesian imitation in interactive computer games. In: Nolfi, S., Baldassarre, G., Calabretta, R., Hallam, J.C.T., Marocco, D., Meyer, J.-A., Miglino, O., Parisi, D. (eds.) SAB 2006. LNCS (LNAI), vol. 4095, pp. 655–666. Springer, Heidelberg (2006)
9. Lasseter, J.: Principles of traditional animation applied to 3d computer animation. SIGGRAPH Comput. Graph. 21(4), 35–44 (1987)
10. Lester, J.C., Stone, B.A.: Increasing believability in animated pedagogical agents. In: Proceedings of the First International Conference on Autonomous Agents, pp. 16–21. ACM (1997)
11. Bryan Loyall, A.: Believable Agents: Building Interactive Personalities. PhD thesis, Carnegie Mellon University (1997)
12. Magerko, B.: Measuring dramatic believability. In: Intelligent Narrative Technologies, pp. 79–82 (2007)
13. Mateas, M.: An oz-centric review of interactive drama and believable agents. In: Veloso, M.M., Wooldridge, M.J. (eds.) Artificial Intelligence Today. LNCS (LNAI), vol. 1600, pp. 297–328. Springer, Heidelberg (1999)
14. Mateas, M., Stern, A.: Façade: An experiment in building a fully-realized interactive drama. In: Game Developers Conference, Game Design Track, vol. 2, p. 82 (2003)
15. McKee, R.: Story: Substance, Structure, Style, and the Principles of Screenwriting. Harper Collins Publishers (October 1997)
16. Ortony, A.: On making believable emotional agents believable. In: Emotions in Humans and Artifacts. MIT Press (2003)
17. Paiva, A., Dias, J., Sobral, D., Aylett, R., Sobreperez, P., Woods, S., Zoll, C., Hall, L.: Caring for agents and agents that care: Building empathic relations with synthetic agents. In: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, vol. 1, pp. 194–201. IEEE Computer Society (2004)
18. Riedl, M.O., Young, R.M.: An objective character believability evaluation procedure for multi-agent story generation systems. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 278–291. Springer, Heidelberg (2005)
19. Thomas, F., Johnston, O.: Disney Animation: The Illusion of Life. Abbeville Press, New York (1981)