An Exploratory Analysis of ECA Characteristics

Adriana Camacho, Alex Rayon, Ivan Gris, and David Novick

Department of Computer Science, The University of Texas at El Paso
500 West University Avenue, El Paso, TX 79968-0518 USA
{accamacho2,amrayon2}@miners.utep.edu, {igris,novick}@utep.edu

Abstract. To help guide design and development of embodied conversational agents, this study reviews the evolving qualities of naturalistic agents by identifying and tracking their most relevant features. The study extends prior taxonomies of characteristics of ECAs, developing a rubric that distinguishes agents' visual and functional characteristics. The study applies the rubric to 15 agents representing different genres of games, distinguishing agents in terms of their naturalistic qualities. The study explores changes in qualities of agents as a function of time.

Keywords: Embodied conversational agent, taxonomy, rubric.

1 Introduction

Embodied conversational agents (ECAs) (Cassell, 2007) have changed the way humans interact with virtual environments and software products. In this paper, to help guide design and development of ECAs, we review the evolving qualities of naturalistic agents by identifying and tracking their most relevant features. In our study, we examined what qualities make a more naturalistic agent by assessing ECAs from video games based on their visual appearance and their functionality. We then assigned weights to the most important capabilities and compared 15 representative ECAs.

The qualities of embodied conversational agents have been analyzed independently (e.g., Pelachaud, 2005; Isbister & Doyle, 2002), but our review did not disclose a comparison of these qualities that developed a comprehensive set of guidelines for evaluation across agents. Existing rubrics include perception domain (context-related), interaction domain (turn-taking), and generation domain (display of expressive synchronized visual and acoustic behaviors) (Pelachaud, 2005). A proposed taxonomy of "Design and Evaluation of Embodied Conversational Agents" suggested measures for describing and evaluating ECAs (Isbister & Doyle, 2002). However, some of the measures were subjective or non-quantifiable, and no agents were evaluated. We propose an improved taxonomy that adds quantitative categories, and we use the taxonomy to evaluate a set of representative agents. Our taxonomy groups the features into two categories, visual and functional. In applying the taxonomy, we weight feature scores to reflect the features' importance and to account for composite features that can only exist in conjunction with others.

T. Bickmore et al. (Eds.): IVA 2014, LNAI 8637, pp. 95–98, 2014.

2 Methodology

We look first at the visual features of ECAs. We identified five key visual features: human likeness, realism, facial expressions, first vs. third person, and motion. Second, we look at the functional features of ECAs. We identified eight key functional features: non-scripted dialogue, verbal communication, level of interaction, group social skills, artificial intelligence level, environment interaction, persona, and nonverbal reaction.

Some agents had been updated across different versions of the games and others remained the same. For example, the agent Navi from the game “The Legend of Zelda” remained constant from its first incarnation at the game’s release in 1998 through its last release in 2012. As different versions of the game were released, the agent’s characteristics (and our assessment of the agent’s realism) did not change. The first version of Navi apparently served its purpose and did not require updates of its appearance or functionality. Other agents were updated. For instance, the agent Cortana from the videogame “Halo” has had several appearance updates that increased her quality as an agent, becoming more realistic in her looks. In considering Cortana, we evaluated both the original and updated versions.

We applied our taxonomy of characteristics to 15 ECAs; the agents selected are presented in Figures 1 and 2. We chose agents based on how representative they were in their specific genre and whether they represented milestones in the game industry when they were introduced. Our evaluation of an ECA was based on our personal experiences of interacting with it and, when an ECA was not available for personal gameplay, on recorded gameplay. The games we were able to play were played in their entirety to provide a full basis for judging agents’ quality. We also considered reviews by game critics and the consensus from players generally. From our 8 key visual and 15 key functional features, we developed a qualities rubric. Agents with scores of 5 for the visual features and 8 for the functional features, for a total of 13 points, would represent the highest-quality, most naturalistic agent.

3 Results

Figure 1 presents a scatter plot for the 15 agents in terms of their visual and functional scores, with the points in plot coded by color for the year the agent was originally released. Darker colors represent more recently released agents. Figure 2 presents a similar scatter plot for the agents in their latest release.

We calculated the mean visual, functional and total scores for the agents’ original and latest versions. Bearing in mind that the evaluated agents do not represent a random sample of the population of commercial ECAs, we note that the agents, on average, improved modestly in visual, functional, and total quality. This improvement, though, came entirely from 3 of the 15 agents: Cortana, Glados, and Milo had a mean increase in score of 0.92 points, while the other agents had no increase in score.

To determine whether the quality of an agent was a function of its year of release, we calculated the correlations between the visual, functional and total scores and the years of original release and latest release. For the agents as originally released, the data suggest that agent quality—visually, functionally, and overall—has increased over time as new agents were developed. That is, newer agents tend to be better agents. But for the agents in their latest versions, there appears to be no significant relationship between quality and time, probably because most recent releases of all but three of the agents were basically the same as their original versions.

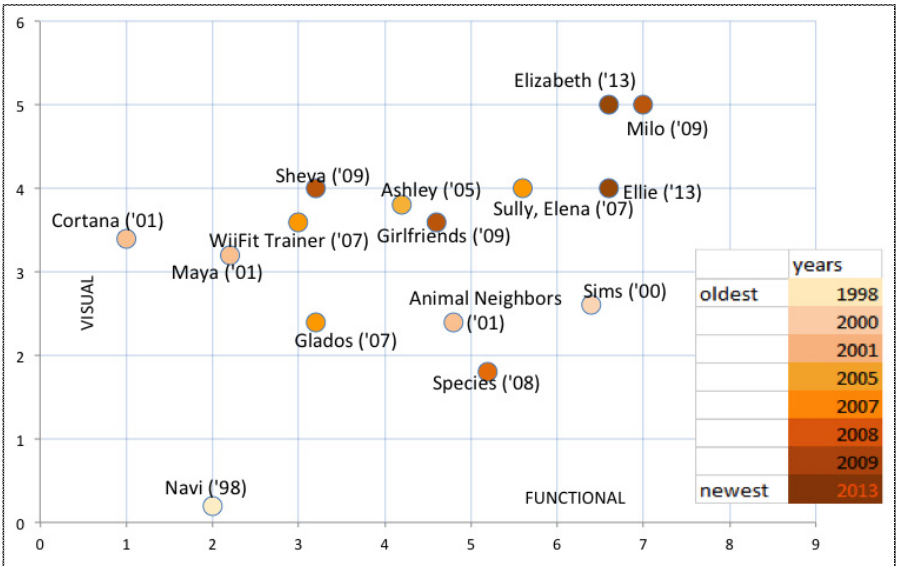

Fig. 1. Visual and functional scores of agents' original version

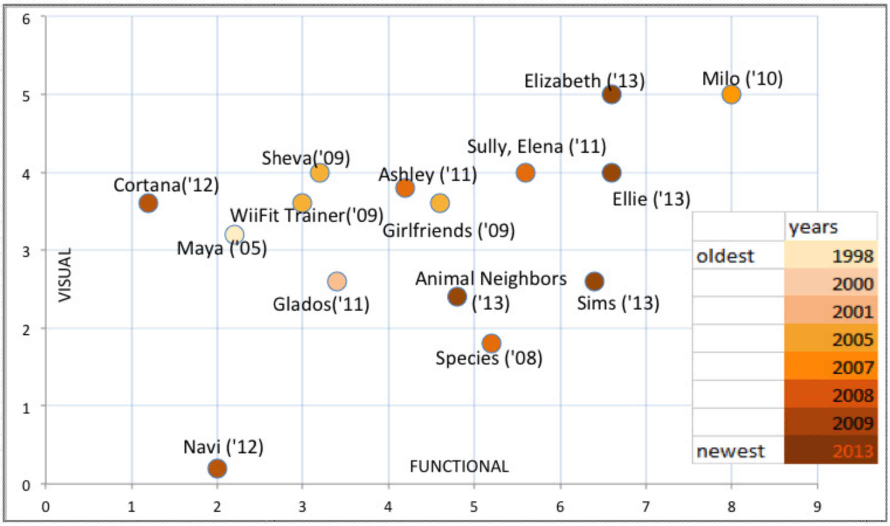

Fig. 2. Visual and functional scores of agents' latest version

Our analysis is subject to several limitations. First, the choice of agents was personal, reflecting the our impressions of game genres and of development milestones for ECAs. The agents we analyzed likely do not represent a true cross-section of ECAs. Indeed, defining such a cross-section would be difficult. A wider range of agents could have presented a fuller picture of characteristics of agents in video games. Second, we did not contact the agents' developers to obtain explanations of how and why they designed the agents with their particular qualities. Third, some agents, such as Sheva and Navi, were apparently developed more for playability than for realism. In a game, a realistically human-like but unusable agent might be interesting to researchers in our field but would be of little interest to actual players.

The rubric, too, has limitations, chief among these that the one-point allocation to each factor is arbitrary. We also tried a version of the rubric that allocated a point to each of the five sub-features in the complex features, but this, even more arbitrarily, gave the complex factors five times more weight than the other factors. A better rubric could be based on a study of users' perceptions that explores the relative importance of the features.

Even with these limitations, the agent-quality rubric could guide development of ECAs not just in videogames but in ECA-based applications more generally. In our current work, we are using the rubric to score and improve the ECAs being developed in our own lab (e.g., Novick & Gris, in press).

References

1. Cassell, J. (ed.): Embodied conversational agents. MIT Press (2000)
2. Isbister, K., Doyle, P.: Design and evaluation of embodied conversational agents: A proposed taxonomy. In: The First International Joint Conference on Autonomous Agents & Multi-Agent Systems (2002)
3. Novick, D., Gris, I.: Building rapport between human and ECA: A pilot study. In: Kurosu, M. (ed.) HCI 2014, Part II. LNCS, vol. 8511, pp. 472–480. Springer, Heidelberg (2014)
4. Parise, S., Kiesler, S., Sproull, L., Waters, K.: Cooperating with life-like interface agents. Computers in Human Behavior 15(2), 123–142 (1999)
5. Pelachaud, C.: Multimodal expressive embodied conversational agents. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, pp. 683–689. ACM (November 2005)
6. Serrels, M.: In Real Life: First Or Third Person - What's Your Perspective? (April 18, 2011), Retrieved from Kotaku: http://www.kotaku.com.au/2011/04/first-or-third-person-whats-your-perspective/
7. Van Vugt, H.C., Hoorn, J.F., Konijn, E.A.: Adapting empathic agents to user experiences. In: AAMAS 2004 Workshop on Empathic Agents, New York (2004)
8. Van Vugt, H.C., Konijn, E.A., Hoorn, J.F., Keur, I., Eliéns, A.: Realism is not all! User engagement with task-related interface characters. Interacting with Computers 19(2), 267–280 (2007)