

# Intelligent and Embodied Agents: A Look into Embodied Conversational Agents

Mohammed Al Orbani

20421873

Bohdan Petrivskyy

20416881

# Introduction

An embodied agent is an autonomous entity that analyzes the environment, and responds accordingly with the purpose of attaining certain goals. More specifically, the principle aim of embodied agent's is to appear relatable to the interactors through both their physical appearance and their responses (Cassell, 2000). For example, a robot that has sensors that guide the robot to make decisions is an embodied agent that analyzes the environment through those sensors and responds accordingly by, for example, moving without colliding with objects. In this paper, the main focus will be embodied conversational agents, that is, agents that are capable of carrying out conversations with other entities, usually humans. That is, a human can interact with the agent and expect a human-like response.

The central problem of creating embodied conversational agents is to be able to make them as realistic as possible. This typically means that the purpose of these agents is to act as humanly as possible. Like conversational agents that are not embodied, a human can expect appropriate (ideally, also correct) from an embodied conversational agent. However, unlike a conversational agent, a Turing Test is neither necessary nor sufficient in this criteria. A Turing Test is a test that a machine passes if it can reliably convince the human interactors that it is a human. The reason why this is not necessary, is that an embodied conversational agent is not trying to hide from the user, it may attempt to be relatable to the user but it is visible and most likely clear that the agent is not a human. The reason why the Turing Test is not sufficient is that the Turing Test does not necessitate the correctness of responses, a machine passes only if the responses resemble a human response regardless of how correct it is. The reason why

correctness is important is that the purpose of typical applications of embodied is to aid a human in performing a certain task, an agent that does not answer correctly is not very useful.

Along with that, since an embodied conversational agent is visible to the user, it introduces new variables, such as facial expressions and gestures, variables outside of the field of NLP, Natural Language Processing.

So, we can see that in order to make an effective embodied conversational agent we must effectively show the following: that the agent produces correct, useful responses and two, that those responses are realistic and relatable to the interactors through both speech/text and expression/movement.

Why is this important? Why would we want to make sure that an embodied conversational agent has these characteristics? Classic human-computer interaction typically involves typing into a program with the intent of receiving something back, for example, the most popular use-case of a computer on the internet is typing a query into Google and retrieving the result (Alexa, 2015). However, this is not a conversation, it is a results-oriented one-way communication of a series of queries followed by the appropriate result.

Conversational agents nowadays are on the rise, with popular products such as Apple's Siri and Google's 'Google Now' because they not only provide a different experience, they provide a seemingly personal experience (Meisel, 2013). An embodied conversational agent is the logical next-step in the evolution of human-computer interaction. There are numerous advantages of the embodiment of a conversational agent, first of all, research reports higher levels of enjoyment, as well as higher levels of recall through embodied user experience (Ijaz, 2011). This

ultimately means that the experience is more memorable and involved than a standard conversational agent or general human-computer interactions. Secondly, the more human a conversational experience is, the higher the believability and trustworthiness of the social interaction (van Mulken, 1988). Imagine you want to sell something without direct human interaction, or even, give important medical advice, it is clear that you must make this experience both memorable and effective for the user. Hence, it is without a doubt very important to study embodied conversational agents and how to implement them effectively by making them both reliable and believable.

# Survey

The main papers surveyed and analyzed were “Fully generated scripted dialogue for embodied agents”, “Embodied Agents: A New Impetus to Humor Research”, “Conversation as a System Framework: Designing Embodied Conversational Agents”, “Enhancing the believability of embodied conversational agents through environment-, self- and interaction-awareness”, “More than just a pretty face: conversation protocols and affordances of embodiment” and “Embodied contextual agent in information delivering applications”. All papers are cited at the end of this paper.

A common vocabulary, feature and gauge of a systems interactivity that is present in almost all of the papers is the concept of “multimodality” and its importance. In the context of embodied agents, a multimodal interface is one that allows multiple methods of interacting with a system as well as multiple ways of conveying a response. For example, a multimodal embodied agent typically accepts both textual and verbal input. More advanced inputs could be things such as human gestures and facial expressions, that is, a machine using a camera would be able to detect and analyze a user’s facial expression and deduce things such as mood or tone. As well, in terms of outputs, the machine can mimic the user’s facial expressions and gestures in order to project the image of personalization and relatability. Other multimodal outputs on top of gestures, movements and facial expressions are the more traditional sound and text outputs.

One paper in particular focused on an interesting feature, integrating humor capabilities in terms of intelligence, emotions and verbal and non-verbal interaction. The main motivation for

this was the adherence to the 'Computers Are Social Actors' paradigm, that is, assigning human characteristics to a computer in order to improve user experience and the quality of interaction. This has some obvious advantages since humor by definition is pleasant experience. However, the suitability of humor must be judged. That is, the culture and age of the user-base must be appropriate and more importantly, the specific task/goal and its seriousness of the embodied conversational agent must first be considered.

Another interesting technique called Fully Generated Script Dialogue that could be used in order to incrementally enhance dialogues generated by embodied conversational agents. This technique involves automated generation of scripts that are then performed to other embodied conversational agents, the process is then tweaked and repeated. That is, we are able to automatically generate dialogue (text, speech and body language) by using 2 embodied conversational agents without the explicit need of a human interactor. The results of this technique could then be used in human-computer interactions. This technique has several advantages, such as being able to create a social computer without the explicit need of heavy interaction with humans. As well, with certain calibrations we can ensure that script is dynamic but also has a guaranteed length, style, and coherence. The major disadvantage is that such a system relies on preprocessing and hence could find certain interactions with people difficult in terms of context, since it would need to know all possible interactions beforehand in order to ensure natural and correct interaction.

On the other hand, another paper talks about the concept of conversation as a framework and using a combination of techniques such as taking timing into consideration, scheduling actions and hardwired reactions. This makes the script not necessarily automated, which could

make it seem less human, however, if calibrated correctly it could actually ensure a consistent and realistic experience as sometimes certain reactions are expected and this system would guarantee it.

Other papers focused on concrete implementations, such as mimicking a real-estate salesperson or a more general information delivering system. A software model for intelligent agents called BDI, Belief-Desire-Intention was used. Based on the programmed environment, goals and how to execute them (belief, desire and intention), the mental state of the system is determined, such as what should the system display and what it shouldn't. Using a Dynamic Belief Network, mental states are injected and emotions are triggered with different durations, intensities etc. depending on the initial belief-desire-intentions of the system. This particular approach will be analyzed and discussed more in the next section as it provides a holistic model of implementing an embodied conversation agent.

# Analysis and Conclusion

All techniques surveyed provide clear advantages and techniques in implementing an effective embodied conversational agent. The BDI technique in particular provides a precise layered model implementation that presents and implements the requirements of creating a believable and expressive embodied conversational agent as outlined by its structure. The general structure contains a “mind” with contextual input and receives and sends information to a Belief Network. Given a goal, dialog input and plans, the next step is to then choose the best plan based on the input and the specific goal of the system. Through this Belief-Desire-Intention system, the embodied conversational agent then incorporates the information into textual, verbal and gestural/emotional output. Supplementing this system with other techniques surveyed in the previous section such as multimodality, timing, humor, generated scripted dialogue, scheduled and hardwired reactions, while considering and working around their disadvantageous, will result in a believable embodied conversational agent.

Overall, it can be seen that there is still a lot of research to be done on embodied conversational agents, especially in the gestural/emotional sense. It is also clear that embodied conversational agents are the next logical successors of conversational agents such as Apple’s Siri and Google’s ‘Google Now’, who they themselves were successors of traditional non-conversational query agents such as Google. This generational view of intelligent agents makes it clear that with every generation, the ultimate goal is to make a more personalized product with the aim of increasing its usability and effectiveness. The advantage that an embodied agent introduces over a (non-embodied) conversational agent is the gestural and emotional



relatability, which in turn was seen as providing a more memorable and believable experience to users. It was also seen that the more natural human characteristics we introduce, the more likely that an embodied conversational agent will succeed in providing a quality Human-Computer Interaction. With this in mind, the different types of features and implementations of effective embodied conversational agents were introduced and explained in order to gain an understanding of not only what makes an embodied conversational agent effective, but how to effectively create one.

## Works Cited

"Alexa Traffic Rank for Google (three month average)". Alexa Internet. Retrieved August 1, 2015.

Beun, Robbert-Jan; de Vos, Eveliene; Witteman, Cilia (2003), "Embodied conversational agents: Effects on memory performance and anthropomorphisation." (PDF), *Proceedings of the International Conference on Intelligent Virtual Agents 2003*, Springer-Verlag, pp. 315–319

Cassell, Justine, et al. "More than just a pretty face: conversational protocols and the affordances of embodiment." *Knowledge-Based Systems* 14.1 (2001): 55-64.

Cassell, Justine, et al. "Designing embodied conversational agents." *Embodied conversational agents* (2000): 29.

Ijaz, Kiran, Anton Bogdanovych, and Simeon Simoff. "Enhancing the believability of embodied conversational agents through environment-, self-and interaction-awareness." *Proceedings of the Thirty-Fourth Australasian Computer Science Conference-Volume 113*. Australian Computer Society, Inc., 2011.

Meisel, William. "The Personal-Assistant Model: Unifying the Technology Experience." *Mobile Speech and Advanced Natural Language Solutions*. Springer New York, 2013. 35-45.

Nijholt, Anton. "Embodied agents: A new impetus to humor research." (2002).

Pelachaud, Catherine, et al. "Embodied contextual agent in information delivering application." *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*. ACM, 2002.

Serenko, Alexander; Bontis, Nick; Detlor, Brian (2007), "End-user adoption of animated interface agents in everyday work applications" (PDF), *Behaviour and Information Technology* **26** (2): 119–132, [doi:10.1080/01449290500260538](https://doi.org/10.1080/01449290500260538).

Van Deemter, Kees, et al. "Fully generated scripted dialogue for embodied agents." *Artificial Intelligence* 172.10 (2008): 1219-1244.

van Mulken, Susanne; André, Elisabeth; Müller, Jochen (1988), "The persona effect: How substantial is it?" (PDF), *HCI '98: Proceedings of HCI on People and Computers XIII*, London, UK: Springer-Verlag, pp. 53–66