Understanding and Designing Gameplay Behaviors for an Embodied Conversational Agent

Max Collins

Carnegie Mellon University

Author Note

Max Collins, Donald Bren School of Information and Computer Sciences, University of California Irvine.

This research was made possible by the CRA-W DREU program and the Articulab at CMU thanks to Dr. Justine Cassell.

For correspondence regarding this paper, please contact Max Collins at mcollin2@uci.edu.

Table of Contents

Abstract **3**

Virtual Peers to Make Learning More Clear: Fostering Curiosity in Children to Drive Self-Initiative **4**

1. Introduction 4

2. Related Work 5

3. Observation of Gameplay Behavior 5

3.1 Gameplay Set Up *6*

3.2 Transcriptions *6*

3.3 Finding Detailed Information *7*

3.4 Categorization of Gameplay Behaviors *8*

3.5 Nonverbal Behaviors *9*

3.6 Calculating Overlap *10*

3. Connecting the Work from Each Team 11

4. Conclusion/Evaluation 12

5. Future Work 12

6. Acknowledgements 13

Works Cited **14**

Abstract

The Articulab's SCIPR (Sensing Curiosity in Play and Responding) project aims to understand how to foster curiosity in children by observing them in small group settings. I am taking part in the development of an embodied conversational agent (ECA), in our study an intelligent virtual child, to engage in a collaborative tabletop game and elicit curiosity during the game play. The virtual child's behavior will be integrated as part of a WoZ (Salber & Coutaz, 1993) system, which allows the above behaviors in a semi-automated matter. To accomplish the overall goal of creating the virtual agent, the SCIPR team split into different roles, and this paper details the work from the perspective of formulating the virtual peer's behaviors, as well as how each role fits together. I personally contributed to the analysis of real human-human interaction to aid the synthesis of a virtual agent's behaviors.

Keywords: Embodied Conversational Agent, Virtual Peer, Curiosity, Gameplay, Learning, Education, Scaffolding.

Virtual Peers to Make Learning More Clear:

Fostering Curiosity in Children to Drive Self-Initiative

1. Introduction

In many sciences, an observational study is a helpful technique that typically involves researchers observing and assessing a group or situation in order to draw inferences without intervening. However, what can you do once you have collected this information? “Data is the new oil” is a term coined by Shivon Zillis (Vanian, 2016). This quote resonates because, like oil, if you don’t know how to use the commodity then it is rendered meaningless. Labs and researchers often struggle to understand how to use and understand data, and how to create meaning from results. The SCIPR project from the Articulab at Carnegie Mellon University is an example of a multifaceted study that employs an observational technique to obtain information from real children interacting in order to interpret and understand their reactions to help create a virtual peer capable of interacting like a real child during gameplay of a tabletop board game. If this virtual peer can be developed and seamlessly interact with real children during gameplay, this allows us to understand and manipulate techniques that highlight curiosity behaviors and scaffolding in children, and potentially elicit curiosity in children purposefully.

For the purpose of this study, the virtual peer will be operated by via the WoZ method, or Wizard of Oz (Salber & Coutaz, 1993). This means that there will be a virtual peer participating live in gameplay, but each action that this virtual peer will perform is chosen by the person operating it from behind the scenes. In a WoZ technique, the virtual system needn’t be self-sufficient, however it must be capable of carrying out all necessary acts in order for operators to instill the illusion of the virtual peer’s natural presence in a given situation.

2. Related Work

Virtual assistants and burgeoning technology of the sort are becoming increasingly ubiquitous in the modern world. In fact, Newark Liberty International, LaGuardia, and John F. Kennedy international airports have embraced this trend and installed Airport Virtual Assistants (Bernstein, 2015). Facebook recently acquired a startup specializing in artificial intelligence to bolster their own virtual assistant (Oreskovic, 2017). While of this is cool and trendy, researchers are digging into the deeper implication of virtual assistants in our lives by aiming to understand how humans interact with them, and how they can be beneficial.

At the Articulab, the SCIPR project's goal is to understand how to elicit curiosity and in turn self-driven motivation to learn in children by means of developing a virtual peer to play a tabletop board game with children. To do so, it is important to understand children's behavior, as well as build off the backs of existing research in virtual assistants, robot-child interactions, and curiosity.

Embodied Conversational Agents, or ECAs, are capable of humanistic behaviors in not just verbal behaviors, but nonverbal behaviors as well (Cassell, 2000). To create an ECA that will be perceived as conversational, one must focus on verbal and nonverbal behaviors both; considering turn-taking heuristics, gaining and releasing attention of others, etc. all shape the user experience.

3. Observation of Gameplay Behavior

In order to go about creating the virtual peer for real children to interact with in the gameplay environment, we must understand how real children interact with one another in the environment that the virtual peer will exist. Six groups of three to four participants per group were recorded while playing the board game for observation purposes.

3.1 Gameplay Set Up

The game that is being used in this study, called "Outbreak," (Alexandra To, 2017) is a Carnegie Mellon University in-house creation, and has rules strategically allowing the participants to have open discussion and use imagination so that the game is not wholly dependent upon luck or competitiveness; this game is also collaborative, in that either all participants win, or all participants lose. It is in each team's best interest to communicate and participate with one another in order to maximize the chances for positive outcome.

Outbreak is a tabletop board game played by up to five players. Players take turns rolling the dice to move a game piece along the board, and they face a collection of ten obstacles along the way. The obstacles, or "rooms" as they are referred to in the game, have a challenge that is unknown to the participants. For each room, participants are given a period of time to ask questions about the room, and these questions are limited in format by question tokens they are given. After this question asking period, participants must discuss and choose up to three "skill cards" (which they receive at the beginning of gameplay) that must match up with the threats in the room that they discover during the question asking phase. If they can correctly cover each threat, then they pass that room.

3.2 Transcriptions

Each of the six gameplay sessions is saved with multiple recordings to capture different video angles, along with voice recordings of each participant. The angles include a webcam view of each participant highlighting their face, as well as an aerial view of each participant capturing the whole table and other participants as well, but facing the targeted participant. In order to do any sort of analysis of the gameplay, it is necessary to understand what happened during gameplay.

For each group, a transcription is manually created based on pre-defined rules of transcription. Utterances beginning/ending time was recorded, along with other noises and noting overlapping speech and simultaneous speech. With each group's transcription, data processing could be done on the full corpus of utterances from the study.

Each transcription is separated into phases in the actual document; that is, each question asking phase and discussion phase of gameplay are marked in order to organize where certain utterances occurred for later examination.

3.3 Finding Detailed Information

With the full corpus transcribed, we focused heavily on two major parts of the Outbreak gameplay: the question asking phase, and the discussion phase. These two phases are where the majority of the utterances and child interaction takes place, so a number of examples from each phase are highlighted as random examples of the phase across all groups. At this point, we have a list of utterances from the question asking phase and the discussion phase that are chosen from all of the groups, and the reason that this list was created instead of focusing on every single utterance from every group will be elucidated by the next step.

This list is deemed representative of the question asking phase and the discussion phase because it spans each group, and it contains examples split evenly among groups and phases. This representative list allows us to initiate further analysis by closely documenting each behavior with the verbal utterance, nonverbal behavior, time stamps in each video, context, precondition, group number, and participant number. With this aggregation of each element of a given behavior, we can understand and analyze trends that will give us a focus of how to handle the corpus as a whole, and in turn the necessary behavioral components for the creation of the virtual peer.

3.4 Categorization of Gameplay Behaviors

Categorization of each behavior involves assessing each behavior while attempting to understand the underlying component. This way, we can work with groups of utterances rather than each utterance as an individual, which in turn allows us to generalize some behaviors and understand under what conditions these behaviors seem to take place. In this case, we are observing gameplay behavior states and curiosity behavior states. Gameplay behaviors are those that take place in gameplay and support the flow of the game. Curiosity behaviors are specifically related to the subject's curiosity and influencing the levels of curiosity during the game. The gameplay behaviors and curiosity behaviors are mapped to one another, and this is so that the end goal of having a functional gameplay reasoner and curiosity reasoner can work in tandem (or independently; the curiosity reasoner intensity can be adjusted independently of the gameplay reasoner being active). The term gameplay reasoner refers to a system to calculate the next logical steps for gameplay, and a curiosity reasoner calculates behaviors depending on a desired influence on participants' curiosity.

The created categories are based on Bales Interaction Process analysis (Fahy, 2015). There are higher-level categories, and lower-level categories within gameplay. The higher-level categories are behaviors that fuel the gameplay and interaction among participants such as suggesting cards to one another, challenging each other's ideas, commenting on events, etc. Lower-level game behaviors are those such as rolling the dice, picking up a card, discarding cards, etc. It is important to take both the high-level and low-level because both must be performed by the virtual agent to make it a capable peer.

A selection of verbal behaviors is chosen from each question asking and discussion phase of each transcription. This tests the strength of the categorized gameplay behaviors, and gives

examples to help build the virtual agent's verbal and nonverbal behaviors. With this information, if a virtual peer must carry out a specific action, there are examples for how the utterance may look based on the data within gameplay behavior states that we have already collected.

3.5 Nonverbal Behaviors

Nonverbal behaviors must be taken into account carefully when developing a virtual agent. This is because real humans will be interacting with this virtual agent, and that interaction must feel natural, and that means that the virtual agent must follow human conversation heuristics and perform appropriate nonverbal behaviors.

For the purpose of this project, we consider nonverbal behaviors with four main components: eye gaze, head gaze, hand position, and facial expression. These four components are observed, and described in detail in order to build correct animations for the virtual peer true to its human counterparts. After all of the nonverbal behaviors from the list of behaviors are detailed, they must be associated with the category that their behavior represents. This way, a nonverbal behavior can be easily and automatically associated with a verbal behavior based on the categorical state. For example, within the "suggest a card" gameplay behavior category, the participants may exhibit specific ways of positioning their head, hands, face, and eyes for a majority of the time, and that would lead us to believe that the virtual peer should act accordingly.

For each state, there is an account of every main component of the nonverbal behaviors. That means that within a specific category, there are counts of how many times a specific behavior occurred (e.g. eye gaze was on another participant, eye gaze was on the table, etc.). This count provides us with great information; we can understand which nonverbal behaviors are

most likely to occur in a given gameplay behavior state, and construct the virtual agent accordingly.

3.6 Calculating Overlap

In each transcription document, we have a chronological series of each utterance throughout the entirety of the session. Associated with each utterance, we have a beginning time and ending time. This information is integral because it allows us to understand when a participant began speaking, and if that was before the previous utterance was completed (i.e. cutting off the previous speaker).

A paper on turn-taking strategies' effect on user evaluation of a virtual agent describes and outlines how turn-taking strategies implemented by virtual agents influence the impression that users have on virtual agents (Ter Maat, Truong, & Heylen, 2010). They find that interrupting a participant results in seeing the agent as more assertive, and leaving pauses between speech is seen as less assertive.

In order to calculate overlap in our study, within a transcription sorted in chronological order I subtracted the beginning time (in milliseconds) from the ending time of the previous utterance. If this number is positive, then the utterance began after the previous one ended, meaning there was no interruption. If this number is negative, then the utterance began before the previous one ended, and there was interruption. The magnitude of the number dictates how far into the previous utterance the cutoff was, or how much of a pause between utterances there was. I also calculated the length of each utterance, the total amount of time each participant spoke within the study, and the percentage of time the participants were speaking while another person was speaking out of their total time speaking.

I did these operations to use my results for beginning to understand when and how frequently participants cut one another off, and in turn begin looking at what implications that could have with perceptions of assertiveness in the conversation.

3. Connecting the Work from Each Team

A strong component to the Articulab's unique savvy is the focus on multidisciplinary work and allowing researchers from different backgrounds to come together and collaborate. To this point, the SCIPR team has a wide variety of talent, in so the project construction is modular at the core. To understand why the work detailed thus far in the paper is necessary, it is important to understand the other pieces of this large project and how everything fits together.

Briefly, another student is working on creating the game description language for the virtual agent, and this is a unique use of GDL because this is an attempt to use it in a collaborative, multi-party game environment. Unlike games like chess where a virtual agent takes a turn, then a human, and so on, in Outbreak there is not so much structure—and each human member can take many courses of action in many different scenarios.

Another student is focusing heavily on the natural language processing component of this project; that is, she is attempting to analyze the corpus of real human behaviors and build a database of behaviors to enable the system to construct sentences during gameplay based on the events occurring at the time.

A third student is constructing the user interface for the human experimenter running the WoZ component of the game during experiments. This is integral to the functioning of the virtual peer because this study relies on the virtual peer appearing as a natural, virtually humanoid companion. The WoZ user interface must be conducive to ease of use so this must be taken into serious consideration.

4. Conclusion/Evaluation

Before leaving campus, the deliverables of highest rank include ensuring that the finite state machine can be reasonably updated, having a list of example behaviors that span each category of gameplay behaviors with detailed information provided per behavior, having defined gameplay behaviors that can be understood and used for annotation, and being able to test gameplay with prototypical systems.

Since all of this will be done by the end of my term here, I would consider this a success; a success because there was good work and lots of learning that occurred on my end, but also because the work that I contributed will help the larger picture and can be built off of in the future by other members of the lab. The Articulab has taught me the importance of working with the expectation that others will eventually need to review the work, and that changed how I document my process for the better.

5. Future Work

Although I will not be a part of the SCIPR team any longer after the conclusion of my internship, I am indeed excited for where this project is headed. The goal intends to impact the way we teach, and should this project infiltrate the education system, that would have an impact on schooling as we know it.

Considering the modularity of the project, the animator will need to update the virtual peer's animations to fit those that have been added since the data analysis that I performed this session. The components that each member of the team worked on will have to be connected and an API must be defined to do so. Once everything is in place technologically, tests will need to be performed to check for reliability and usability, and then the system can be updated as the SCIPR team sees fit based off of interactions with more groups of child participants.

6. Acknowledgements

I would like to thank Dr. Justine Cassell for taking on interns through the DREU program this year, because that is the core to why any of this was possible for me. I would also like to thank the CRA-W DREU program for sponsoring me and allowing me to participate in this great program. Lastly, the entire team at the Articulab for being so great to work with and so phenomenal at what they do. Namely, I'd like to thank Dr. Zhen Bai for supervising me during my work with the SCIPR project and teaching me crucial practices of research that I will continue to practice.

Works Cited

Alexandra To, J. H. (2017, July). Modeling and Designing for Key Elements of Curiosity: Risking Failure, Valuing Questions. *Digital Games Research Association, 14*, 18.

Bernstein, P. (2015, March 30). *Forget About Siri. Meet Ava, Your New Personal Assistant*. Retrieved August 1, 2017, from IndieWire: http://www.indiewire.com/2015/03/forget-about-siri-meet-ava-your-new-personal-assistant-63644/

Fahy, P. J. (2015, September 9). Online and Face-to-Face Group Interaction Processes Compared Using Bales' Interaction Process Analysis (IPA).

Oreskovic, A. (2017, July 31). *Facebook bought an AI startup that could turn its middling virtual assistant into a Siri killer*. Retrieved August 2, 2017, from Business Insider: http://www.businessinsider.com/facebook-acquires-ozlo-ai-startup-2017-7

Vanian, J. (2016, July 11). *Why Data Is The New Oil*. Retrieved July 31, 2017, from Fortune: http://fortune.com/2016/07/11/data-oil-brainstorm-tech/

Cassell, J. (2000). Embodied conversational interface agents. *Communications of the ACM*, *43*(4), 70–78. https://doi.org/10.1145/332051.332075

Salber, D., & Coutaz, J. (1993). Applying the Wizard of Oz technique to the study of multimodal systems. In *Proceedings of EWHCI* (pp. 219–230). https://doi.org/10.1007/3-540-57433-6_51

SCIPR. (n.d.). Retrieved August, 2017, from http://articulab.hcii.cs.cmu.edu/projects/scipr/

Ter Maat, M., Truong, K. P., & Heylen, D. (2010). How turn-taking strategies influence users' impressions of an agent. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6356 LNAI*, 441–453.

https://doi.org/10.1007/978-3-642-15892-6_48