Kashihara, A. et al. (Eds.) (2024). Proceedings of the 32nd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Optimization of Non-Verbal Information for English Conversation Agents Using Interactive Evolutionary Computation

Yuma SHIMOSAKA[a*], Emmanuel AYEDOUN[b] & Masataka TOKUMARU[b]
[a]*Graduate School of Science and Engineering, Kansai University, Japan*
[b] *Faculty of Engineering Science, Kansai University, Japan*
*k307597@kansai-u.ac.jp

Abstract: As English becomes increasingly important globally, many agent-based conversation practice environments struggle to maintain learner motivation due to a lack of personalized behavior. This study proposes optimizing a conversational agent's non-verbal cues—such as nodding and voice characteristics—through interactive evolutionary computation to enhance learners' motivation. Participants engaged in role-play scenarios across eight settings, providing feedback after each interaction. The agent's behavior was iteratively optimized, and approximately 90% of participants reported increased willingness to interact, suggesting that personalizing non-verbal behavior can significantly improve motivation in language learners.

Keywords: English learning, Interactive Evolutionary Computation (IEC), Non-verbal Information, Learning Motivation

1. Introduction

In recent years, the importance of English conversation skills has grown due to globalization, leading to the rise of online platforms using conversational agents—software that interacts with users in natural language. These agents, enhanced by advancements in natural language processing, are valuable tools for educational support.

To assess their effectiveness, it is important to consider psychological factors. Jean Piaget's theory emphasizes the interaction between emotions and intelligence in learning (Decarie, 1978). Effective learning relies on the continuous interaction of these elements. Additionally, mutual trust and interaction between educators and learners are crucial (Hawkins et al., 1987; Bickmore et al., 2005), with non-verbal cues like gestures and facial expressions being essential for information transmission (Fox, 1993).

Educational conversational agents (Pedagogical Conversational Agents: PCA) aim to replicate human-like interactions in one-on-one learning (Baylor et al., 2005; Gulz, 2005). Embodied Conversational Agents (ECA) and Animated Pedagogical Agents (APA) use non-verbal cues to enhance user trust and familiarity (Johnson et al., 2000). Non-verbal information in human-robot communication helps mitigate task challenges and improve performance (Admoni et al., 2016).

Despite its importance, designing universally appealing English conversation agents remains challenging due to diverse user preferences. This study uses Genetic Algorithms (GA) (Holland, 1992) and Interactive Evolutionary Computation (IEC) (Takagi, 2001) to optimize non-verbal cues based on user preferences and evaluate if this improves learning motivation.

2. Proposed System

2.1 System Outline

To test whether optimizing the non-verbal cues expressed by English conversation agents can enhance users' motivation to learn, we designed a system where the agents exhibit various non-verbal behaviors, such as facial expressions, voice tones, and nods during conversations. By optimizing the agent's behavior, we believe users will feel more familiar with the agent, increasing their motivation to speak English.

The virtual avatar used in this study, referred to as an English conversation agent, is illustrated in Figure 1. The study aims to determine if optimizing the agent's non-verbal cues based on user preferences contributes to their motivation to learn English through engagement with the agent. Figure 2 shows an overview of the system, which operates in two phases: the English conversation phase, where the user and agent engage in dialogue based on a given scenario, and the optimization phase, where the system refines the agent's non-verbal cues.

2.2 English Conversation Phase

This system uses role-playing in English conversation scenarios, where students practice by acting out roles in predefined situations. It includes 80 scenarios, each lasting around 30 seconds, set in one of eight locations: a cafe, hotel, airport, movie theater, clothing store, library, hospital, or road intersection, with ten scenarios per location.

The user's name is "Jan," and the agent is "Becky," allowing them to address each other by name. After completing a conversation in one scene, the location and background change. Table 1 shows the background image and scenario for a cafe setting.

2.3 Optimization phase

The system evaluates English conversations between users and agents that display non-verbal cues, optimizing agents to match user preferences. Users rate each agent on a scale from 1 to 10.

Figure 3 illustrates the optimization process. Initially, candidate solutions are generated randomly and presented to users. Users then engage in conversations with these candidates and rate them. After one generation, genetic algorithms (GA) process the ratings, and new agents are generated and presented to users. This process repeats for several generations.

Figure 1. English Conversation Agent

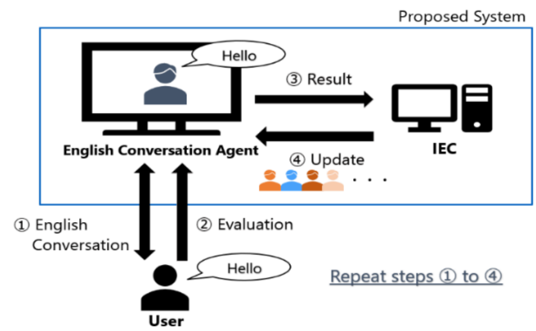

Figure 2. Interaction Flow

Table 1. Cafe background images and example scenarios

Background Image	Scenario
	User : Hello, I'd like to order a chicken sandwich and an Americano, please. Agent : Sure thing! What toppings would you like on the sandwich? User : Lettuce and tomato, please. Also, I'll add a milkshake to that. Agent : Of course. I'll bring that right over for you.

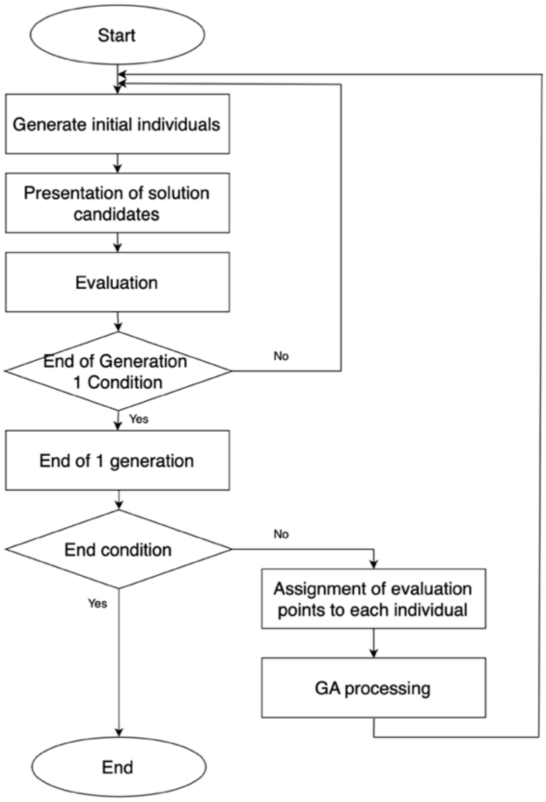

Figure 3. Flowchart of the Interactive Evolution Computation

Table 2. Genetic representation of nonverbal information expressed by the agent

Depth of nod	10
Speed of nod	10
Timing of nod	10
Number of nods per cycle	4
Pitch Speed	10
Facial expression when talking	10

2.4 English Conversation Agent in Action

The English conversation agent in this system performs three actions: "idling," "talking," and "listening." In the "idling" state, it blinks naturally. When talking, it moves its mouth and hands, and when listening, it nods to show attention. The system optimizes the agent's nonverbal cues in the "idling" state.

Nonverbal cues are encoded as genes for optimization using a genetic algorithm (GA). Table 2 outlines these genes: six genes representing different aspects of communication, including nod depth, speed, timing, and frequency for listening, and pitch speed and facial expressions for speaking. Gene values range from 0 to 3 for nod depth and 0 to 9 for others, resulting in 400,000($10^5 * 4$) possible combinations.

2.5 User Interface

The interface is structured in three distinct phases: pre-conversation, conversation, and evaluation. In the initial phase, as depicted in Figure 4, the display presents the English conversation scenario, generation count, individual identifier, and an initiation button. The conversation phase transitions the interface to a bilingual display, featuring both English and Japanese text (Figure 5). During the evaluation phase, illustrated in Figure 6, the interface prompts the user to input an assessment score and activate a "Next Lesson" button. Post-evaluation, this button transforms into a "Next Generation" option, facilitating the creation of subsequent agent generations. This procedural cycle persists for the predetermined number of generational iterations.

3. Experiment

This experiment was approved by the Research Ethics Review Committee of Organization for Research and Development of Innovative Science and Technology (ORDIST) in Kansai University, Japan.

Figure 4. Pre-Conversation Interface

Figure 5. Interface during Conversation

Figure 6. Evaluation Interface

Table 3. *GA parameters*

Population size	8 individuals
Number of generations	10 generations
Gene length	6
Method of mating	Uniform crossover
Selection method	Roulette wheel + Elite conservation
Mutation rate	10%

3.1 Experiment Overview

In this experiment, participants engaged in role-play English conversations with virtual agents and rated each agent on a scale from 1 to 10, with 1 being the best and 10 the worst. After conversing with all agents, they interacted with two agents in sequence: one optimized for their preferences (Agent A) and one rated highest in the first round (Agent B). To control for order effects, participants were split into two groups: one group interacted with Agent A first, followed by Agent B, and the other group did the reverse. Participants then completed questionnaires comparing Agent A and Agent B, as well as their overall experience with the system.

The study involved 12 university students in their 20s, all native Japanese speakers. The genetic algorithm used 8 individuals per generation over 10 generations. The final questionnaire, shown in Table 4, used a 5-point scale to compare the two agents. Next, the questionnaire items regarding the experience of using the system are shown in Table 5. This questionnaire consisted of four questions, along with space for free-form responses.

Table 4. Questionnaire items regarding comparison of English conversation agents

Number	Question
Q1	Which was your preferred agent?
Q2	Which agent did you feel more motivated to learn by using?

Table 5. Questionnaire items regarding the experience of using the experiment

Number	Question
Q3	Did you feel that the agent has been optimized to your liking over the generations?
Q4	Have more agents become more approachable with each generation?
Q5	Which nonverbal information influenced the first impression?
Q6	Did you feel an evaluation burden?

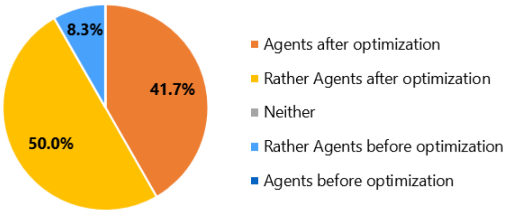

Figure 7. Q1 Results

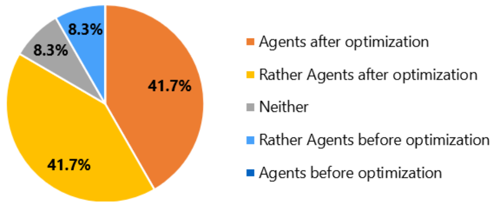

Figure 8. Q2 Results

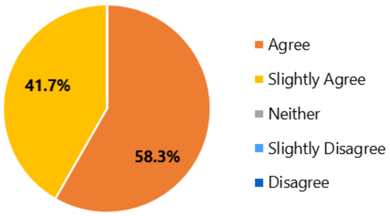

Figure 9. Q3 Results

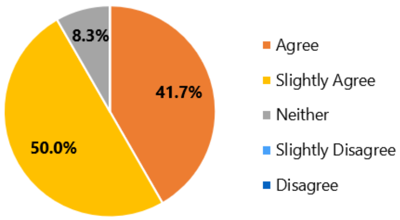

Figure 10. Q4 Results

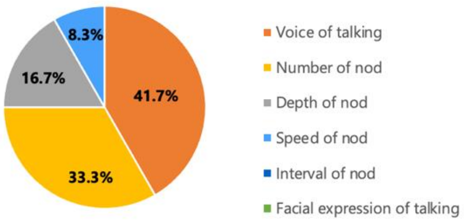

Figure 11. Q5 Results

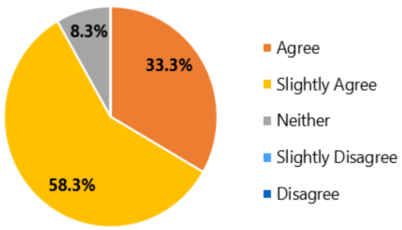

Figure 12. Q6 Results

3.2 Results

The post-experiment questionnaire results show that optimizing the English conversation agent's nonverbal cues increased participants' motivation to learn English. Q1 indicates a preference for the optimized agent, while Q2 shows that learning motivation was higher with the optimized agent compared to before optimization. Results from Q3 and Q4 suggest that participants noticed more optimized and familiar agents with each generation, but Q6 highlights that the evaluation process was burdensome for them.

Participants' comments included observations such as varying sentence difficulty and the impact of changing backgrounds on engagement. Some noted that nonverbal cues, like nodding, significantly affected their impressions, while others found it difficult to understand facial expressions and suggested that more nonverbal information would be helpful.

4. Discussion

First, let’s discuss the results of the questionnaire comparing English conversation agents. As shown in Figure 7, the optimized English conversation agents were generally preferred over the agents that were highly rated before optimization. Figure 8 further illustrates that optimizing the nonverbal cues of the English conversation agents enhances the participants' motivation to learn. However, one participant rated both Q1 and Q2 higher for the pre-optimized agent. Analysis revealed that the pre-optimized agent’s gene sequence was similar to that after optimization, suggesting that the initial random sequence matched the participant’s preferences from the start.

Next, we discuss the results of the post-experiment questionnaire regarding the experience of using the system. The results shown in Figure 9 indicate that the system successfully optimizes the agents in line with the participants' preferences. Figure 10 suggests that optimizing the nonverbal cues improves the familiarity of the agent, and there may be a significant relationship between the robot's familiarity and its nonverbal information.

Figure 11 shows that the most common aspect participants focused on was the "speaking voice" of the agent, followed by the "number of nods." This suggests that the frequency of nodding is an important factor in English conversation, potentially influencing the participants' confidence in the interaction.

Figure 12 reveals that many participants found the evaluation process burdensome, likely due to the need to assess multiple agents and focus on both the English text and the agent’s nonverbal cues.

Free responses included comments like "very difficult" and "challenging." Participants also noted scenario-related challenges, such as "I didn’t have time to observe the agent during difficult scenarios" and "There were differences in the difficulty level of the scenarios." This implies that when the scenario's difficulty is high, even an agent aligned with the participant's preferences might receive a lower evaluation. However, varying the scenario content is necessary to maintain participants' interest.

5. Conclusion

In this study, we developed a system that optimizes an English conversation agent’s nonverbal cues based on user preferences using Interactive Evolutionary Computation (IEC). Our goal was to assess whether enhancing the agent’s impression could boost user motivation for English conversation. Participants engaged in role-plays with the agent and evaluated its performance.

Results showed that optimizing the agent's nonverbal cues did increase participants' motivation to practice English. However, this boost in motivation is not solely due to the optimization of nonverbal cues, as other factors like interaction strategies also play a role. Additionally, the current system places a significant burden on users during the optimization process. Future work will focus on refining interaction strategies and minimizing the user evaluation burden to create a more user-friendly and effective learning environment.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers JP21K12099 and JP22K18011.

References

Agenta Gulz. (2005). Social enrichment by virtual characters – differ entail benefits. Journal of Computer Assisted Learning. 21(6). 405-418.

Amy Baylor, Yanghee Kim. (2005). Simulating instructional roles through pedagogical agents. International Journal of Artificial Intelligence in Education. 15(2). 95-115.

Barbara Fox. (1993). The human tutorial dialog project. Hillsdale. NJ: Erlbaum.

Henny Admoni, Thomas Weng, Bradley Hayes, Brian Scassellati. (2016). Robot nonverbal behavior improves task performance in difficult collaborations. Proc. 11th ACM/IEEE Int. Conf. Human–Robot Interact. (HRI). 51-58.

Hideyuki Takagi. (2001). Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation. Proceedings of the IEEE. 89(9). 1275-1296.

Jan Hawkins, Roy Pea. (1987). Tools for bridging the cultures of everyday and scientific thinking. Journal of Research in Science Teaching. 24(4). 291-307.

John Henry Holland. (1992). Adaptation in Natural and Artificial Systems. MIT Press.

Lewis Johnson, Jeff Rickel, James Lester. (2000). Animated pedagogical agents: Face- to-face interaction in interactive learning environments. International Journal of Artificial Intelligence in Education. 11(1). 47-78.

Therese Gouin Decarie. (1978). Affect development and cognition in a Piagetian context. The development of affect. Spring er. Boston. MA. 1. 183-204.

Timothy Bickmore, Rosalind Wright Picard. (2005). Establishing and maintaining long- term human-computer relationships. ACM Transactions on Computer-Human-Interaction. 12(2). 293-327.