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

Exploring the Benefits of Strategic Hesitations in Language Learning Robots

Ryusei AZUMA[a*], Emmanuel AYEDOUN[b] & Masataka TOKUMARU[ab]
[a]*Graduate School of Science and Engineering, Kansai University, Japan*
[b]*Faculty of Engineering Science, Kansai University, Japan*
*k403626@kansai-u.ac.jp

Abstract: This study investigates the impact of introducing conversational pauses and self-adaptor gestures in robots to enhance human-likeness during English conversation practice. As globalization increases the demand for English proficiency, there is increasing interest in using conversational robots for language practice. However, conventional robots often lack natural disfluencies, making interactions feel overly artificial, especially for beginners. This study focuses on enhancing the robot's non-verbal behaviors by implementing conversational pauses and self-adaptor gestures such as fidgeting at predetermined intervals during scripted conversational scenarios. The experiment involved 22 men and women in their 20s. An experimental evaluation revealed that approximately 90% of participants perceived an increased sense of human-likeness when the robot exhibited these pausing and hesitation behaviors. These results suggest that strategically incorporating naturalistic disfluencies and human-like self-adaptor gestures into a conversational robot's non-verbal repertoire can significantly increase its perceived anthropomorphism and potentially improve user engagement in language learning contexts.

Keywords: L2 Conversation Practice, Non-Verbal Behaviors, Human-Robot Interaction, Simulated Hesitation

1. Introduction

The increasing globalization has led to a growing interest in English conversation practice. However, since effective practice requires at least two people, there is a significant challenge in creating regular opportunities for learners (Nishino, 2008). The issue is particularly pronounced for beginners, who often experience anxiety and resistance when practicing with others. In response to these challenges, there has been a notable trend towards utilizing communication robots as practice partners (Engwall et al., 2021).

The effectiveness of conversation practice heavily relies on repetition, which allows learners to internalize sentence construction in their non-native language. Therefore, it is crucial to design robots that can maintain and enhance users' motivation to engage in conversation (Hjalmarsson et al., 2007). Improving a robot's human-likeness can significantly contribute to improving motivation for conversation (Złotowski et al., 2015).

Previous studies have attempted to enhance the human-likeness of robots by reproducing human-like behaviors. For instance, Salem et al. (2011) demonstrated that incorporating hand and arm gestures led to more positive user evaluations. Wigdor et al. (2016) found that introducing gestures as fillers before robot responses improved both the perceived human-likeness and likability of the robot.

However, there is a notable gap in the existing research regarding the robot's gestures during ongoing conversation. Human conversations naturally include pauses, regardless of the speaker's proficiency. In contrast, conventional robots typically deliver predetermined text monotonously, maintaining consistently high fluency. This presents an opportunity to enhance a robot's human-likeness by mimicking the pauses and disfluencies of human speech.

This study aims to explore whether introducing hesitation gestures during pauses can enhance the perceived human-likeness and friendliness of conversational robots. We hypothesize that by incorporating these human-like disfluencies, users will recognize that “even a robot that usually speaks fluently sometimes hesitates” potentially increasing the robot’s perceived friendliness and naturalness. In this context, we define “Gesture Behavior” as gestures accompanied by speech that combine both verbal and non-verbal fillers. Our research seeks to evaluate the impact of these behaviors on user perception and engagement in the context of English conversation practice with robots.

2. Proposed System

2.1 Robot Overview and Behaviors Design

In this study, we utilize NAO as our communication platform (NAO -- ALDEBARAN, 2022). We implemented hesitation behaviors during conversations. These behaviors are designed to simulate the robot contemplating its response to the next statement. We developed three types, as illustrated in Figure 1 (a-c). Each behavior lasts for 3 seconds and is accompanied by verbal expressions of uncertainty. We define “hesitation levels” to differentiate between the three behavior types. To ensure consistency, the robot performs these behaviors at predetermined times during an “extroverted conversation” - active communication primarily focused on information exchange with the user. The robot’s speech is paused to create conversational gaps, during which the behaviors are performed as fillers.

To prevent user discomfort from a stationary talking robot, we implemented an automatic gesture generation system, a built-in NAO function that determines appropriate gestures based on speech content.

(a) (b) (c)

Figure 1. Robot’s hesitation behavior levels (a) Hesitation level 1, (b) Hesitation level 2, and (c) Hesitation level 3.

2.2 Combining Conversation with Hesitation Behavior

To control the timing of the robot’s behaviors, we prepared scenarios and conversation flows. Table 1 lists the topics executed in the scenarios, while Figure 2 presents the conversation flowchart.

We integrated a natural language processing system based on GPT-3.5 Turbo from the Azure OpenAI Service (Azure OpenAI Service models, 2024). For speech-to-text functionality, we employ OpenAI’s Whisper base model and Azure OpenAI Service’s Whisper (Whisper, 2023; Azure OpenAI Service models, 2024).

Table 2 presents an example conversation demonstrating how hesitation behaviors are embedded within the dialogue. Each behavior is assigned a hesitation level, with the level for responses to user questions varying based on the duration of the user’s question (T), as defined in equations (1) to (3).

$$0.0 \leqq T < 2.0 \quad (1)$$

$$2.0 \leqq T < 4.0 \quad (2)$$

$$4.0 \leqq T \quad (3)$$

Table 1. *Topics Covered in Conversation Scenarios*

Scenario1	Scenario2
Favorite food	Favorite drink
Favorite animal	Favorite sport
Places you would like to visit in Japan	Places you would like to visit overseas

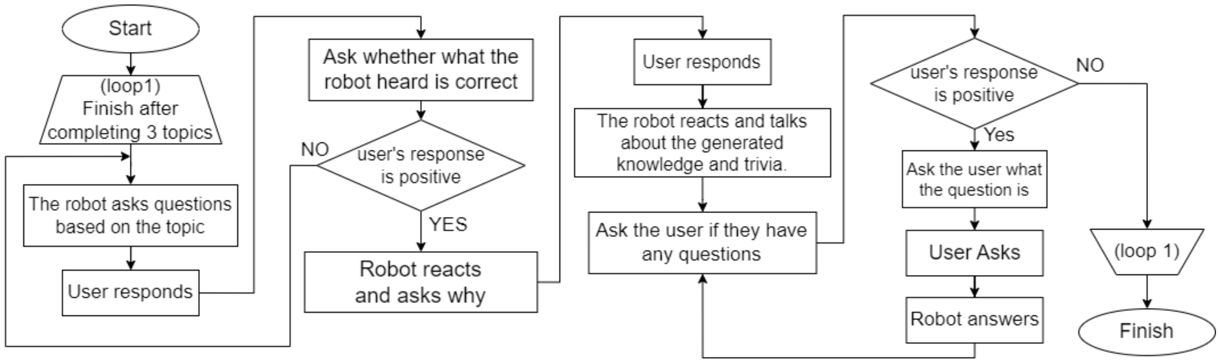

Figure 2. Conversation Flowchart

Table 2. *Example Conversation on The Topic "Favorite food"*

Robot	User
What is your favorite food?	
	My favorite food is orange.
OK. Is orange your favorite food?	
	Yes.
Oh, that's interesting. **(Hesitation behavior: level2)** Why do you like it?	
	Because, I like its taste.
I see. That's a good reason. **(Hesitation behavior: level 3)** *(Robot say about trivia about orange)* Do you have any question about orange?	
	Yes.
Oh, what is your question?	
	What is it famous for?
Nice question. **(Hesitation behavior: level 1~3)** *(Answer to questions)* Do you have any other question about orange?	

3. Experimental Evaluation

3.1 Methodology

The experiment involved one-on-one conversations between subjects and two types of robots: a robot equipped with a comparison model (without behaviors) and a robot equipped with the proposed model (with behaviors). Subjects interacted with each robot for approximately 15 minutes. To account for order effects of both model and scenario, subjects were divided into four groups, as shown in Table 3. The experiment involved 22 men and women in their 20s.

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.

Table 3. *Experimental Group Assignments for Model and Scenario Order*

Group	First	Second
1	Scenario 1 – Proposed	Scenario 2 – Comparison
2	Scenario 1 – Comparison	Scenario 2 – Proposed
3	Scenario 2 – Proposed	Scenario 1 – Comparison
4	Scenario 2 – Comparison	Scenario 1 – Proposed

3.2 Questionnaires

The evaluation process involved two questionnaires. The first, assessing impressions of individual robots, comprised two questions (shown in Table 4) and free-form comments. Participants rated each question on a five-point Likert scale: “Strongly Agree / Agree / Neutral / Disagree / Strongly Disagree”. The second questionnaire, comparing the two robot types, consisted of four questions (shown in Table 5) and free-form comments. This comparative assessment used a five-point scale: “Strongly Prefer Proposed Model / Prefer Proposed Model / No Preference / Prefer Comparison Model / Strongly Prefer Comparison Model”.

Table 4. *Questionnaire Items for Individual Robot Evaluation*

Q1	To what extent did you perceive the robot as capable of expressing emotions?
Q2	How natural did your communication with the robot feel?

Table 5. *Questionnaire Items for Comparative Robot Evaluation*

Q3	Which robot exhibited more human-like characteristics?
Q4	With which robot did you feel a greater sense of familiarity?
Q5	Which robot provided a more natural communication experience?
Q6	Which robot more effectively motivated you to engage in English conversation?

3.3 Result

The effectiveness of the proposed robot in expressing human-likeness and enhancing communication was evaluated through analysis of questionnaire responses. Data were collected from two sets of questionnaires: one administered after each experimental interaction (Figures 3 and 4) and a comparative questionnaire conducted after all experiments (Figures 5 to 8). All numerical values are rounded to one decimal place for consistency.

3.3.1 Individual Robot Impressions

Q1 showed that 68.2% of participants perceived that the robot equipped with the proposed model exhibited emotions (Figure 3). Q2 showed that 86.4% of participants reported that the robot with the proposed model communicated naturally (Figure 4).

3.3.2 Comparative Robot Evaluation

The proposed model consistently outperformed the comparison model in expressing human-likeness and friendliness (Q3 and Q4, Figures 5 and 6). Regarding naturalness of communication (Q5), 81.8% of participants found the proposed model’s communication style more natural and human-like (Figure 7). Interestingly, for motivation to speak English (Q6), results show a noticeable edge for the comparison model (Figure 8).

3.3.3 Qualitative Feedback Analysis

Free-form comments offered additional insights into participants' perceptions. For the proposed model, participants appreciated the robot's "thinking time," variety in speaking speeds, and the ease of conversation due to natural pauses. However, some participants found the "hesitation movements" excessive and noted a lack of correlation between speech content and the duration of hesitations.

The feedback also revealed an influence of participant proficiency on robot preference. Some participants suggested that the proposed model might be more suitable for less proficient English speakers, while the comparison model's smooth conversation flow appealed to more advanced users.

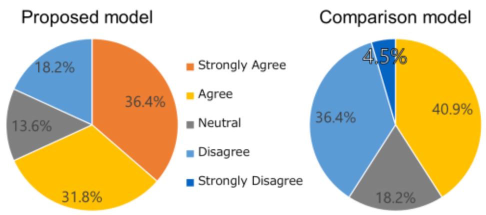

Figure 3. Result of Q1

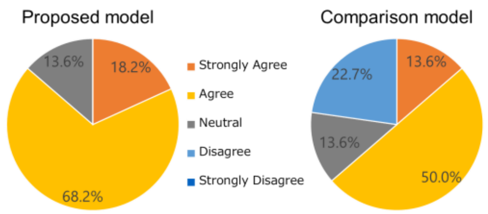

Figure 4. Result of Q2

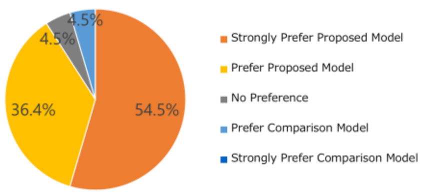

Figure 5. Result of Q3

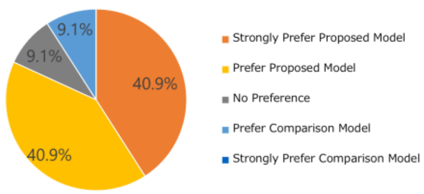

Figure 6. Result of Q4

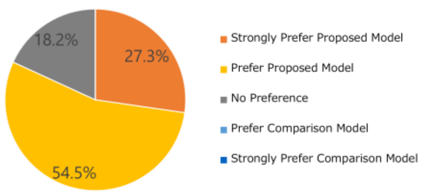

Figure 7. Result of Q5

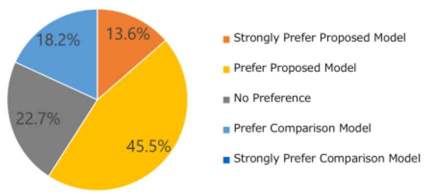

Figure 8. Result of Q6

4. Discussion

Figure 3 suggests that hesitation gesture behaviors successfully replicated aspects of human conversation. The proposed model's ability to reproduce natural communication is further supported by the comparative data in Figure 7, which shows a clear preference for the proposed model over the comparison model in terms of communication naturalness.

The proposed model also demonstrated superior performance in expressing human-likeness (Figure 5) and improving the robot's friendliness (Figure 6). These findings suggest that the implemented behaviors effectively simulated human-like characteristics, enhancing both the perceived human-likeness and friendliness of the robot.

In addition, participant feedback highlighted the positive impact of conversational pauses introduced by the hesitation gesture behaviors. These pauses were reported to contribute to the robot's perceived fluency, making interactions feel more natural. This observation supports the hypothesis that natural human-like communication can be reproduced through the strategic implementation of hesitation gesture behaviors.

However, this study revealed potential areas for refinement. Some participants reported discomfort with the duration and frequency of the behaviors, as evidenced by Figure 8 and qualitative feedback. This discomfort may be attributed to the fixed duration and high frequency of the behaviors. The lack of correlation between speech content and hesitation time was also noted as a point of concern.

An interesting finding revealed that less proficient English speakers preferred the proposed model for its human-like qualities, while more advanced users favored the comparison model for smoother conversations. This observation indicates that user proficiency may be a crucial factor in determining the effectiveness of different robot interaction styles in language learning contexts.

These findings enhance our understanding of human-robot interaction in educational settings, especially for language learning. While the proposed model shows promising results in enhancing perceived naturalness and human-likeness, future work should focus on refining the implementation of hesitation behaviors to better align with user expectations and proficiency levels. Additionally, investigating adaptive systems that can adjust behavior frequency and duration based on user characteristics could further enhance the effectiveness of robot-assisted language learning.

5. Conclusion

This study proposed and evaluated an English conversation robot that performs hesitation gesture behaviors to enhance human-likeness and friendliness. In a comparative experiment, participants conversed with two robots—one with the proposed model and one without—following set scenarios and conversation flow. Participants then rated the robots using questionnaires.

The results demonstrated that the robot with the proposed model successfully reproduced more natural communication, significantly improving perceived human-likeness and friendliness. However, the study also revealed important areas for refinement. Some participants expressed discomfort with the duration and frequency of the robot's hesitation behaviors. Additionally, we observed that participants' evaluations varied depending on their English conversation proficiency, suggesting a need for adaptability in the robot's behavior.

Future work will focus on developing a more sophisticated model that correlates the robot's speech content and information density with its hesitation time. We also aim to implement an adaptive system that adjusts the frequency of hesitation behaviors to match the user's proficiency level. These improvements should further enhance the effectiveness of robot-assisted language learning.

References

Azure OpenAI Service models. (2024). Microsoft. https://learn.microsoft.com/en-us/azure/ai-services /openai/concepts/models

Engwall, O., Lopes, J., & Åhlund, A. (2021). Robot interaction styles for conversation practice in second language learning. *International Journal of Social Robotics*, *13*(2), 251-276.

Hjalmarsson, A., Wik, P., & Brusk, J. (2007). Dealing with DEAL: a dialogue system for conversation training. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, 132-135.

NAO – ALDEBARAN. (2022). Aldebaran & United Robotics Group. https://www.aldebaran.com/en/nao

Nishino, T. (2008). Japanese secondary school teachers' beliefs and practices regarding communicative language teaching: An exploratory survey. *Jalt Journal*, *30*(1), 27-50.

Salem, M., Rohlfing, K., Kopp, S., & Joublin, F. (2011). A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction. *2011 RO-MAN*, 20th IEEE International Symposium on Robot and Human Interactive Communication, 247-252

Whisper. (2023). GitHub. https://github.com/openai/whisper

Wigdor, N., de Greeff, J., Looije, R., & Neerincx, M. A. (2016). How to improve human-robot interaction with Conversational Fillers. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN),* 219-224.

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: opportunities and challenges in human–robot interaction. *International journal of social robotics*, *7*, 347-360.