Timing Matters? Response Delays in Human-Robot Interaction with Misty.

*Note: Sub-titles are not captured in Xplore and should not be used

Antonia Sunna
KTH Royal Institute of Technology
Stockholm, Sweden
asunna@kth.se

Aryan Dolas
KTH Royal Institute of Technology
Stockholm, Sweden
aryand@kth.se

Jesús Gerardo Ortega Peimbert
KTH Royal Institute of Technology
Stockholm, Sweden
jgop@kth.se

Maria-Alexandra Florea
KTH Royal Institute of Technology
Stockholm, Sweden
maflorea@kth.se

Tafarrel Firhannoza Pramono
KTH Royal Institute of Technology
Stockholm, Sweden
tafarrel@kth.se

***Abstract*—This study explores the impact of response delays in backchanneling and gaze on human-robot interaction (HRI) using Misty, a social robot. The participants engaged in an interactive animal-guessing game while Misty responded with immediate or delayed responses. Visual engagement was manually measured through video analysis and the likeability and perceived intelligence ratings of the participants were collected through post-interaction questionnaires. Although the results did not show statistically significant differences in engagement time, likeability, or perceived intelligence between delay and no delay conditions, qualitative feedback provided valuable insights. Some participants highlighted delays as a source of uncertainty, but also appreciated Misty's natural and responsive communication style. Observations indicated changes in participant focus from Misty to task-related materials following negative responses, reflecting changes in engagement patterns. These findings highlight the importance of timing and social cues in HRI design, contributing to the development of more effective and engaging robotic systems.**

***Index Terms*—Human-robot interaction, backchanneling, response delays, gaze, engagement, game, social cues, social robotics.**

I. Introduction

Research on robot communication has consistently demonstrated the significant benefits of non-verbal cues, such as gaze and backchanneling, in enhancing human-robot interaction (HRI). These signals are crucial in making interactions feel more natural, authentic, and responsive, mirroring the conversational dynamics of human-human communication. Non-verbal cues also influence how humans perceive robots, impacting engagement, likeability, and trustworthiness.

Backchanneling refers to verbal and non-verbal cues used by a listener to signal attentiveness, understanding, or agreement without interrupting the speaker's turn. Examples include verbal signals like "uh-huh," "I see," or "got it," as well as non-verbal behaviors such as nodding, smiling, facial expressions, and the use of eye contact[1].

In HRI, backchanneling behaviors are essential social cues that significantly improve interaction quality. For instance, robots employing backchanneling are perceived as more engaged and better team members in collaborative human-robot tasks[2]. In child-robot interactions, backchanneling increases children's engagement, leading to greater attention toward the robot and more interactive speech [3]. However, the perception of robots is not uniform and may depend on individual human traits.

A study by [4] found that backchanneling behaviors, such as gaze and head nodding, influenced likeability differently for introverts and extraverts. Interestingly, while backchanneling can enhance engagement, it may paradoxically lead to robots being perceived as less competent [2]. Reciprocal behavior also affects likeability, as pure reciprocal strategies in game-based interactions are generally preferred over combined strategies [5]. These findings underscore the complexity of designing effective backchanneling models for robots.

Gaze, defined as the direction and focus of a person's eye movements, is another critical component of non-verbal communication. In human interactions, mutual eye contact enhances feelings of connection and helps regulate conversational turn-taking [6]. Gaze also facilitates attention regulation, social signaling (e.g., showing interest or disengagement), and information gathering, which aids in interpreting environmental and social cues.

In HRI, shared attention through gaze cues has been shown to improve interaction quality. For instance, during handovers, shared gaze leads to faster human responses and a better perception of interaction quality [7]. Similarly, human-human interactions benefit from gaze cues, enhancing performance in cooperative tasks [8]. Both gaze and backchanneling are crucial for grounding mutual understanding during task-oriented

Identify applicable funding agency here. If none, delete this.

instructions, with their effectiveness varying depending on motivation and instructional context [9].

Despite these parallels, significant differences exist between human-human and human-robot backchanneling behaviors. For example, backchanneling in robots often requires tailoring to specific emotional or contextual dynamics, as shown by [10]. Incorporating such cues into robot behaviors can improve the flow, efficacy, and naturalness of interactions.

Non-verbal cues also contribute to how "human" robots are perceived. Robotic anthropomorphism, or the design of human-like robots, can positively influence the relationship between human personality traits and robot likeability in social settings [11]. However, excessive anthropomorphism risks triggering the uncanny valley effect [12], where robots appear eerily human-like, evoking discomfort or distrust. This highlights the delicate balance required in designing anthropomorphic robots that engage effectively without becoming unsettling.

These findings highlight the importance of integrating non-verbal cues like backchanneling and gaze into robot behaviors to enhance interaction quality, engagement, and perceived likeability.

II. Related Work and Research Gap

While many studies have investigated backchanneling in Human-Robot Interaction (HRI), research focusing on **delays in robot backchanneling remains scarce and has mixed results**. Delays allow users to process robot cues, such as gaze shifts or backchanneling behaviors, which may influence the perception of the robot's attentiveness and responsiveness. Delays in robot responses can enhance awareness and compliance in interactions, particularly in gaze-based communication [13]. Delays have also been found to increase user frustration, anger and arousal while decreasing satisfaction, as measured through facial recognition and electrodermal activity [14]. However, more targeted research is needed to clarify the impact of delays in robot communication.

Guy Hoffman has particularly looked at fluency in robot collaboration, and found that timing plays a role in a range of human-robot interaction scenarios. Humans are sensitive to timing and interaction fluency as it is central in spoken dialogue, turn-taking, interruptions and hesitations that can affect both efficiency and emotional response [15][16][17]. Moreover, non-verbal behavior such as gestures, gaze and other social communication cues are sensitive to timing [15][18][19]. Fluency is typically also associated with the robot being perceived as more reliable, capable and having enhanced trust and appreciation from human collaborators. Smooth and well-timed interactions reduce misunderstandings and interruptions, leading to better task outcomes. It also impacts how naturally humans can integrate robots into their workflows, especially in collaborative settings [20].

A study in [21] explored how human backchanneling is affected when robots use invitational cues, such as pauses or gestures. They found that humans are more likely to backchannel in the robot interaction when pauses and gestures were used, indicating that these invite the human to talk. While this study highlighted the importance of timing in promoting human responses, there is limited understanding of how delayed robot backchannels influence these reciprocal dynamics, particularly in comparison to real-time interactions.

Another pilot study by [22], showed no significant difference in engagement between delays and no delays in robot movements during a turn-taking game. These findings suggest that not all delays might have a negative effect or disruptions in participant engagement. However, participants in this study looked at the robot's face more when they experienced the interaction as less fluent, which might indicate the participants were looking for social cues to understand the short disruption of the game.

Another study by [23] explored the effects of movement delays on perceptions of fluency, trust, and anthropomorphism in a pilot study. While no significant effects were found due to limited statistical power, the study underscores the importance of even short hesitations in shaping user perceptions. It also highlights the need for more robust study designs and refined measures to assess embodied and emotional responses to delays in robot interactions.

[24] found that delays encouraged participants to think more deeply during interactions. Instead of focusing on the robot's face, they paid more attention to the board, hand movements, and problem-solving. This led to better gameplay, with participants trying and comparing strategies to win. Delays also changed the participants behavior, such as longer focus on questions, brief eye contact, and some frustration with wrong answers, showing increased cognitive effort.

This study aims to address a gap in the research on backchanneling in human-robot communication, particularly focusing on the impact of delays. While backchanneling has been examined in terms of its effects on likeability and perceived intelligence, there is limited evidence on how delays in backchanneling influence these measures. Existing studies on delays largely center on robot motor actions during tasks like turn-taking, leaving backchanneling delays underexplored. By investigating delays in social cues, this study seeks to advance our understanding of how such delays affect user perceptions and interactions. Ultimately, this work aims to contribute to the development of affect-aware robotic systems capable of managing time delays and responding effectively to user emotional states in human-robot interactions.

III. Method

A. *Hypothesis*

This project examines how response delays in backchanneling and gaze influence a robot's likeability, perceived intelligence, and visual engagement, using a game-based interaction with a companion robot named Misty [25]. We hypothesize that:

H1. Delays will increase the participants' visual engagement. A lack of immediate response may cause participants to focus their attention on the robot, attempting to understand the reason for the delay [24]. This increased attention may also extend to observing the test leader for cues or additional context, heightening overall visual engagement within the interaction.

H2. Delays will have a negative impact on the likeability. Delays in response could frustrate participants or disrupt the natural flow of interaction [20], making the robot appear less engaging or personable. Such interruptions in the conversational rhythm may reduce participants' perception of the robot's sociability, leading to lower likeability ratings.

H3. Delays will have a negative impact on perceived intelligence. We hypothesize that delays might be interpreted as an indicator of limited cognitive processing or inefficiency, leading participants to perceive the robot as less intelligent or competent [2]. In human interactions, prompt responses are often associated with competence, and a lack of immediacy could undermine the robot's perceived intellectual capabilities.

B. Experimental Design

This study will use an interactive animal-guessing game, What am I?, to examine how specific social behaviors, like backchanneling, gaze patterns, and response delays, affect participants' likeability ratings of a social robot and their visual engagement during interactions. The experimental design enables a controlled investigation into the role of social cues (backchanneling and gaze) and timing (delayed vs. immediate responses) in shaping participant perceptions and engagement.

1) Robot and Interaction: Participants interact with Misty, a social robot, by asking binary questions to identify animal images displayed on a table. Misty responds based on predefined behavioral scripts that vary across experimental conditions. The interaction is designed to simulate natural conversational behaviors and test their impact on user experience.

2) Wizard of Oz Approach: To ensure precise control of Misty's behaviors, a *'Wizard of Oz'* methodology is employed. An operator controls Misty's responses using pre-recorded audio files to simulate specific backchanneling behaviors and introduce response delays as necessary. The approach also facilitated the logging of Misty's answers, with detailed timestamps recorded for each interaction organized based on participant IDs.

3) Gaze Behavior: Misty's gaze system incorporates face-tracking technology to alternate between making eye contact with participants and looking at the images on the table. Gaze behavior is kept consistent with both conditions to assess its influence on participant engagement and perception.

4) Conditions (The independent variable): The study employs a between-group design with the following conditions:

- **Group 1**: No response delays (immediate answers)
- **Group 2**: Response delays (introduced to simulate processing or cognitive effort)

The delays were implemented through a boolean flag system in the software controlling the robot. Both conditions included natural short delays corresponding to the human operator's response time. For Group 1, only these natural operator delays were present. For Group 2, an additional fixed 2-second artificial delay was programmatically inserted before each robot response, simulating extended processing time or cognitive effort. This implementation was consistent across all response types (yes/no/maybe answers) throughout the interaction.

5) Measures (The dependent variables):

1) **Visual Engagement** Participants' visual engagement during the experiment was analyzed using a gaze tracking system. The system evaluated whether participants were visually focused on the robot by monitoring head orientation and eye openness relative to the robot's position. **Key metrics included:**

- The percentage of time participants spent looking at the robot (engagement percentage).
- The average duration of each gaze.
- The total number of times participants looked at the robot.

2) **Likeability and Perceived Intelligence Ratings** Participants complete post-interaction questionnaires to provide self-reported ratings of Misty's likeability and perceived intelligence, using all items for both parameters in accordance with the Godspeed Questionnaire [26]. By systematically manipulating response timing and monitoring participant reactions, this design aims to explore the interplay of social cues and temporal dynamics in the interaction with Misty.

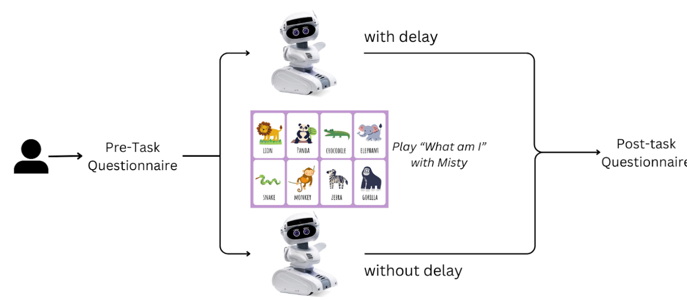

Fig. 1. Experiment Design Diagram

C. Technical Design

The system is developed using HTML powered by a Flask server to create a graphical user interface (GUI) for Wizard-of-Oz experiments. The Flask server communicates with the robot via its API and streams live camera feeds during the experiment using the WebSocket protocol.

The Gaze Detection module uses computer vision algorithms to track participants' faces and estimate head orientation and eye openness. This information is processed in real time to determine the participants' gaze direction and engagement status. The pipeline outputs engagement data, which is used for both system monitoring and post-experiment analysis.

Python threading is utilized to handle concurrent tasks, such as streaming the camera feed to the GUI and processing visual data for gaze analysis. The GUI enables the operator to monitor and control interactions in real-time, such as enabling/disabling delays in the robot's responses and allowing the operator to respond to participant questions.

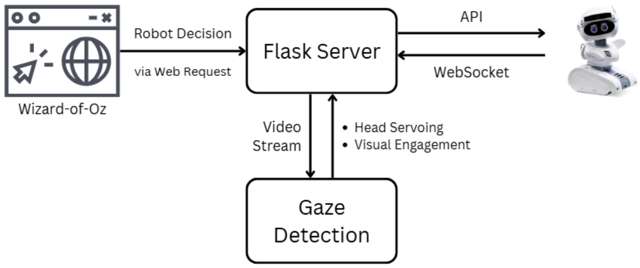

Fig. 2. Technical Design Diagram

D. Procedure

Participants provided informed consent before participating in the study. They also shared demographic details, including age, gender, and first language. The test leader then provided instructions about the task, ensuring that participants fully understood the procedure. Participants then completed five rounds of an interactive guessing game, where their goal was to identify five animals from a printed list of 40. The list was displayed between the participant and Misty. Participants were encouraged to ask Misty questions, which could be answered with 'Yes,' 'No,' or 'Maybe.' In each round, the participants could ask up to five questions before making their final guess. Upon completing the final round, participants evaluated Misty by filling out a post-experiment survey. This survey included 10 items from the Godspeed Questionnaire, with five questions on likeability and five on perceived intelligence. If participants had any additional questions for the instructors at any point, they were encouraged to ask, and their inquiries were addressed promptly.

IV. Results

A. Participants

A total of 10 participants (8 females, 2 males) were recruited through online student groups, ensuring a level of control over the experiment by selecting individuals with an inherent interest in the study. Participants ranged in age from 21 to 37 years (M = 28.9, SD = 5.69). Inclusion criteria required participants to have no prior interaction with the robot used in the study to ensure unbiased engagement. All participants provided informed consent prior to their involvement in the study. No monetary or material compensation was offered; participants were thanked for their time and effort.

B. Analysis (statistical methods)

Two tests were conducted for evaluating visual engagement. One for the visual engagement rate and one for the visual engagement time. The visual engagement time is the actual time in seconds the participants spent looking at Misty. The visual engagement rate is the time in percentage; the time participants spent looking at Misty divided by the total time of the experiment.

1) Visual Engagement Rate: A Welch's independent samples t-test was conducted to compare **engagement rates** between the *Delay and No Delay* conditions. The results indicated no significant difference, t(7.39) = -0.80, p = .446. The mean engagement rate for the Delay condition was 18.88 (SD = 12.98), with a median of 19.60 and a range of 6.40 to 37.30. For the No Delay condition, the mean engagement rate was 24.70 (SD = 9.67), with a median of 24.40 and a range of 15.70 to 36.60. The 95% confidence interval for the mean difference was [-22.76, 11.12], which includes 0, indicating no evidence of a statistically significant difference.

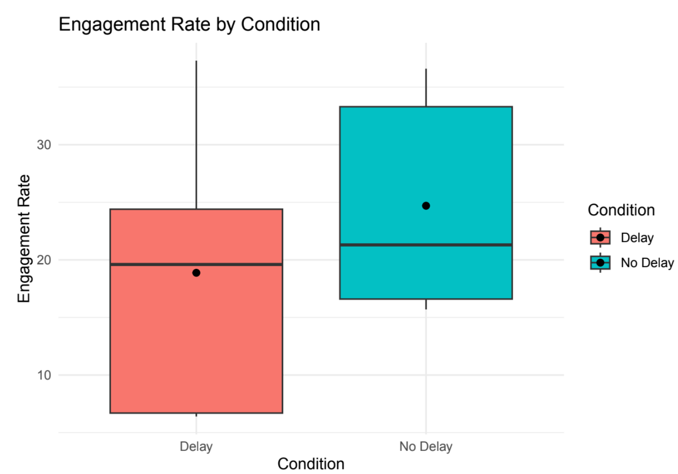

Fig. 3. The boxplot visualize the median (bold black bar), mean (black dot), and confidence intervals for the engagement rate.

TABLE I
STATISTIC TABLE OF ENGAGEMENT RATE

Statistic (Engagement Rate %)	Delay	No Delay
Mean	18.88	24.70
Median	19.60	24.240
Standard Deviation (SD)	12.98	9.67
Range (Min - Max)	6.40-37.30	15.70-36.60
95% Confidence Interval	[-22.76, 11.12]	[-22.76, 11.12]
p-value	0.4465	0.4465

2) Visual Engagement Time: A Welch's independent samples t-test was conducted to compare **engagement time** between the Delay and No Delay conditions. The results indicated that there was no significant difference in engagement time between the two conditions, t(5.77) = -0.93, p = .388. The mean engagement time for the Delay condition was 83.8 seconds (SD = 44.87), compared to 104.6 seconds (SD = 21.66) for the No Delay condition. The 95% confidence interval for the mean difference was [-75.86, 34.26], which includes 0, suggesting no evidence of a significant difference.

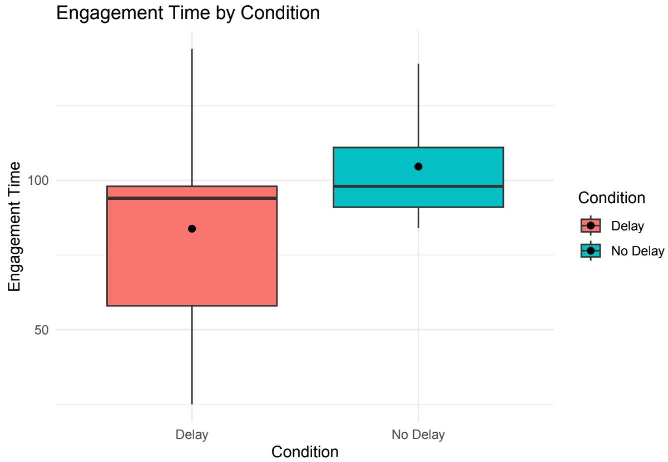

Fig. 4. The boxplot visualize the median (bold black bar), mean (black dot), and confidence intervals for the engagement time.

TABLE II
STATISTIC TABLE OF ENGAGEMENT TIME

Statistic (Engagement Time)	**Delay**	**No Delay**
Mean	83.8	104.6
Median	85.75	96.00
Standard Deviation (SD)	44.8687	21.66333
Range (Min - Max)	25-144	94-111
95% Confidence Interval	[-75.86, 34.26]	[-75.86, 34.26]
p-value	0.388	0.388

3) Likeability: A Welch's independent samples t-test was conducted to compare **likeability ratings** between the Delay and No Delay conditions. The results indicated no significant difference in likeability, $t(7.90) = 0.62$, $p = .554$. The mean likeability rating for the Delay condition was 22.4 ($SD = 0.71$), compared to 21.4 ($SD = 0.61$) for the No Delay condition. The 95% confidence interval for the mean difference was [-2.74, 4.74], which includes 0, indicating no evidence of a significant difference.

TABLE III
STATISTIC TABLE OF LIKEABILITY

Statistic (Likeability)	**Delay**	**No Delay**
Mean	4.48	4.28
Median	5	4
Standard Deviation (SD)	0.7141428	0.6137318
Range (Min - Max)	3-5	3-5
95% Confidence Interval	[-0.18, 0.58]	[-0.18, 0.58]
p-value	0.2937	0.2937

4) Perceived Intelligence: A Welch's independent samples t-test was conducted to compare **perceived intelligence ratings** between the Delay and No Delay conditions. The results indicated no significant difference in perceived intelligence, $t(8.00) = 0.72$, $p = .492$. The mean perceived intelligence rating for the Delay condition was 19.2 ($SD = 1.03$), compared to 17.0 ($SD = 1.00$) for the No Delay condition. The 95% confidence interval for the mean difference was [-4.85, 9.25], which includes 0, suggesting no evidence of a significant difference.

Fig. 5. The boxplot visualize the median (bold black bar), mean (black dot), and confidence intervals for the likeability questionnaire scores of the participants.

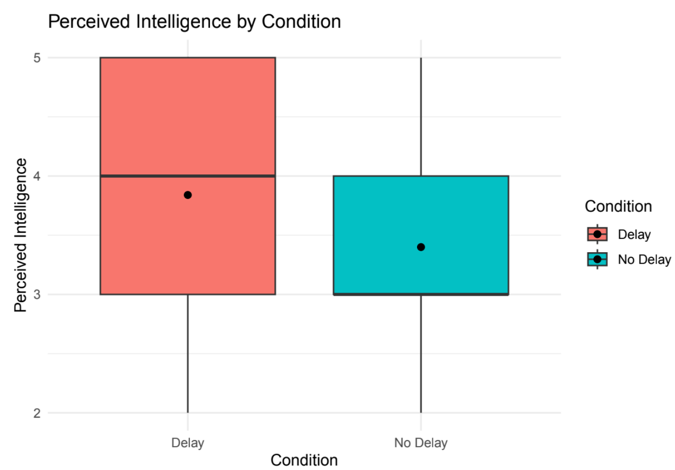

Fig. 6. The boxplot visualize the median (bold black bar), mean (black dot), and confidence intervals for the perceived intelligence questionnaire scores of the participants.

TABLE IV
STATISTIC TABLE OF PERCEIVED INTELLIGENCE

Statistic (Intelligence)	**Delay**	**No Delay**
Mean	3.84	3.4
Median	4	3
Standard Deviation (SD)	1.027943	1.00
Range (Min - Max)	2-5	2-5
95% Confidence Interval	[-0.14, 1.02]	[-0.14, 1.02]
p-value	0.1316	0.1316

V. DISCUSSION

The findings indicate no statistically significant differences between the delay and no-delay conditions across the three measured outcomes: visual engagement, likeability, and perceived intelligence. The results of the t-tests suggest that the null hypothesis cannot be rejected for any of these variables.

This lack of significance implies that the presence or absence of response delays did not meaningfully influence participants' visual attention to the robot, their perceived likeability of the robot, or their evaluation of its intelligence. These findings may suggest that response timing, within the ranges tested, is not a critical factor in shaping participant engagement or perceptions in this specific context.

A. General Discussion

Observations During Data Collection and After the Experiment

1) Focused on Paper: During the data collection it was noted that some participants seemed really focused with the task and spent a lot of time during the experiment with looking at the animals instead of interacting with the robot. One common behavior was that the participant looked at the robot when directly asking the question or only looked at Misty when she was giving an answer.

Our collected data suggests that after receiving a negative answer from Misty to their question, most participants shifted their focus more toward the paper, slowly avoiding eye contact with Misty. For example, Participant ID 10 displayed this pattern: at the beginning of the interaction, the durations of eye contact were relatively long—around 4–5 seconds. However, as the session progressed, the durations of eye contact decreased significantly, often lasting only 1–2 seconds. This pattern suggests that the participant was increasingly disengaged from Misty and redirected their focus toward analyzing the papers with the animals list for clues, trying to win.

A similar pattern was observed during rounds where participants failed to correctly guess the animal. In these cases, participants avoided eye contact even when Misty was speaking or responding to their questions. This contrasts with the beginning of the game, where participants often maintained focus on Misty while asking questions or awaiting her responses.

However, this behavior was not universal. For instance, Participant ID 4 provided feedback that they felt Misty gave answers much faster when they maintained eye contact with her. This perception seemed to influence their behaviour, as they sustained eye contact with Misty for longer periods. In fact, Participant ID 4 had the longest engagement time among all participants—144 seconds—with an engagement rate of 37.3%. This feedback suggests that for some participants, maintaining eye contact with the robot may have created a sense of improved responsiveness and interaction efficiency.

For example, Participant ID 4 guessed 4 animals out of 5, while Participant ID 10 guessed only 1 animal out of 5. This difference in success rate could reflect the contrasting engagement styles and focus observed between the two participants.

2) System Inconsistency: Due to limited robot computational power (explained in V-B), some delays in the no-delay condition happened sporadically. This might have decreased potential differences between the groups, although since it happened only once or twice for one participant, it is not likely this effect has made a huge impact. It is possible that intermittent responses could potentially enhance the focus on the robot, although it is hard to know for sure without further studies on this particular matter.

3) Participant Feedback: Several participants commented on the enjoyable aspects of interacting with Misty. Participant ID 4 remarked that the experiment was "nice" and particularly appreciated Misty's responsiveness when maintaining eye contact, which they felt made the interaction feel faster and more natural. This suggests that Misty's design successfully created moments of perceived responsiveness, which could enhance user satisfaction. Similarly, Participant ID 8 shared that they found Misty's kind way of speaking to be pleasant, even speculating on Misty's ability to understand accents, which indicates trust in Misty's social communication abilities. Participant ID 6 also found the experiment engaging, noting their enjoyment in asking a variety of questions to test the quality of Misty's responses. Participant ID 9's feedback, describing the experiment as "perfect," reflects a very positive experience without concerns.

While the feedback was generally positive, participants also highlighted some challenges during their interaction with Misty. A recurring theme that we noticed was related to timing and clarity in Misty's responses. Participant ID 1 (with delay) noted that Misty was "slow sometimes", which was echoed by Participant ID 6 (without delay), who commented on delays in Misty's reactions (for example, maybe was confused with no), leading to moments of uncertainty. Participant ID 10 found it unclear when to speak and when not to, which could indicate that the turn-taking dynamics were not entirely intuitive. These points suggest that improvements in response timing and clearer conversational cues could enhance the naturalness of the interaction.

Participants also pointed out issues with Misty's communication style. For example, Participant ID 7 described Misty's phrasing of "ah no no no" as unnatural and suggested using a different prompt for such responses. Some participants highlighted difficulties related to the experimental task itself. Participant ID 5 shared that they often forgot the questions they had asked previously, which impacted the quality of their guesses. Participant ID 10 also found it hard to understand some of Misty's answers, which could point to the need for clearer or more consistent verbal responses.

These findings based on the participant feedback align with the study's focus on exploring engagement and interaction quality, providing valuable insights for further development.

B. ***Limitations***

1) Homogeneity: One key limitation of this study is the homogeneity of the participant sample. The majority of participants were recruited from KTH, consisting primarily of students and personnel. This demographic concentration may introduce bias and limit the generalizability of the findings to broader, more diverse populations. Additionally, the small sample size poses a constraint on the study's statistical power, making it challenging to detect subtle effects or draw robust general conclusions. Future research would benefit from larger and more diverse participant pools to enhance the reliability and applicability of the results.

2) Wizard of Oz: Another limitation of this study lies in the use of the Wizard-of-Oz methodology. Since the robot was controlled by a human operator, the response times were partially influenced by the operator's actions. This introduces the possibility that some participants in the no-delay condition experienced occasional delays due to human factors. However, the Wizard-of-Oz process was consistent across both conditions, with the delay condition incorporating the same human response times plus the preprogrammed additional delays. Therefore, it is unlikely that any sporadic delays in the no-delay condition significantly impacted the interaction or the study's results.

3) Robot Computational Power: During individual feature testing, all components, including gaze detection, head servoing, and command execution from the GUI, performed smoothly without issues. However, after integrating all the features, we noticed a decline in the overall system performance. Specifically, the video streaming has slower frame rates and noticeable lag when Misty processes a command involving head gestures. Moreover, the commands sent from the GUI were not always successfully delivered to Misty. This is likely due to API request conflicts between the head servoing operations and the GUI commands. As a result, the gaze detection module output cannot be used for analysis. We decided to measure the visual engagement manually by reviewing the recorded videos for each participant. During this process, we documented the specific timestamps of visual engagement in a file for each participant. This data was then used to calculate both the total engagement time and the engagement rate.

VI. Conclusion

This study provides valuable insights into the role of response timing, gaze behavior, and social cues in human-robot interaction (HRI). While the results showed no statistically significant differences between the delay and no-delay conditions across visual engagement, likeability, and perceived intelligence, qualitative and observational data reveal important findings,

A. ***Future Studies and Future Work***

Future directions could be comparing facial expressions with machine learning (ML) and comparing this with the likeability. How many times and for how long does the participants have certain facial expressions, and why?

Another potential study could be to do the same experiment again without the Wizard-of-Oz. Instead, it could be interesting to use pre-programmed large language model (LLM) responses. In that case, we could know for sure that the human response time of the wizard does not affect the visual engagement, likeability or perceived intelligence scores.

It would also be interesting to see how intermittent delay responses could impact the human-robot interaction.

Based on user feedback, future work could focus on enhancing Misty's naturalness and clarity, such as replacing unnatural phrases with more human-like responses and a broader variety of them. Another possible future improvement could be incorporating reminders for previously asked questions and turn-taking features to reduce user confusion. Additionally, exploring personalized interaction models and refining non-verbal cues that could further improve Misty's responsiveness and likeability rates from users.

References

[1] Wikipedia, *Backchannel (linguistics)*, Retrieved December 6, 2024, from https://en.wikipedia.org/wiki/Backchannel_(linguistics), 2024.

[2] M. F. Jung, J. J. Lee, N. DePalma, S. O. Adalgeirsson, P. J. Hinds, and C. Breazeal, "Engaging robots: Easing complex human-robot teamwork using backchanneling," in *Proceedings of the 2013 Conference on Computer-Supported Cooperative Work And Social*, https://robots.media.mit.edu/wp-content/uploads/sites/7/2015/01/Jung2013.pdf, 2013, pp. 1555–1566.

[3] H. W. Park, M. Gelsomini, J. J. Lee, T. Zhu, and C. Breazeal, "Backchannel opportunity prediction for social robot listeners," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2017, pp. 2308–2314.

[4] C. Thepsoonthorn, K. I. Ogawa, and Y. Miyake, "The relationship between robot's nonverbal behaviour and human's likability based on human's personality," *Scientific reports*, vol. 8, no. 1, p. 8435, 2018.

[5] E. B. Sandoval, J. Brandstetter, U. Yalcin, and C. Bartneck, “Robot likeability and reciprocity in human-robot interaction: Using ultimatum game to investigate the effect of robot’s likeability on human’s reciprocity,” *International Journal of Social Robotics*, vol. 13, pp. 851–862, 2021. DOI: 10.1007/s12369-020-00658-5.

[6] M. Belkaid, K. Kompatsiari, D. De Tommaso, I. Zablith, and A. Wykowska, “Mutual gaze with a robot affects human neural activity,” *Science Robotics*, vol. 6, no. 58, 2021. DOI: 10.1126/scirobotics.abc5044.

[7] A. Moon, D. M. Troniak, B. Gleeson, *et al.*, “Meet me where i’m gazing: How shared attention gaze affects human-robot handover timing,” in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 2014, pp. 334–341.

[8] J.-D. Boucher, U. Pattacini, A. Lelong, G. Bailly, F. Elisei, S. Fagel, *et al.*, “I reach faster when i see you look: Gaze effects in human–human and human–robot face-to-face cooperation,” *Frontiers in neurorobotics*, vol. 6, p. 3, 2012.

[9] J. Oppenheim, J. Huang, I. Won, and C. M. Huang, “Mental synchronization in human task demonstration: Implications for robot teaching and learning,” in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 470–474.

[10] P. Shahverdi, K. Rousso, J. Klotz, I. Bakhoda, M. Zribi, and W. Y. Louie, “Emotionally specific backchanneling in social human-robot interaction and human-human interaction,” in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2023, pp. 4059–4064.

[11] A. S. Arora, M. Fleming, A. Arora, V. Taras, and J. Xu, “Finding ̌201cȟ201d in hri: Examining human personality traits, robotic anthropomorphism, and robot likeability in human-robot interaction,” *International Journal of Intelligent Information Technologies (IJIIT)*, vol. 17, no. 1, pp. 1–20, 2021.

[12] M. Mori, K. F. MacDorman, and N. Kageki, “The uncanny valley [from the field],” *IEEE Robotics & automation magazine*, vol. 19, no. 2, pp. 98–100, 2012.

[13] H. Admoni, A. Dragan, S. S. Srinivasa, and B. Scassellati, “Deliberate delays during robot-to-human handovers improve compliance with gaze communication,” in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 2014, pp. 49–56.

[14] E. Yang and M. C. Dorneich, “The effect of time delay on emotion, arousal, and satisfaction in human-robot interaction,” in *Proceedings of the human factors and ergonomics society annual meeting*, Sage CA: Los Angeles, CA: SAGE Publications, vol. 59, 2015, pp. 443–447.

[15] G. Hoffman, M. Cakmak, and C. Chao, “Timing in human-robot interaction,” in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 2014, pp. 509–510.

[16] C. Chao and A. L. Thomaz, “Timing in multimodal turn-taking interactions: Control and analysis using timed petri nets,” *Journal of Human-Robot Interaction*, vol. 1, no. 1, pp. 4–25, 2012.

[17] Y. Okuno, “Providing route directions: Design of robot’s utterance, gesture, and timing,” in *Proceeding of the ACM/IEEE international conference on Human-robot interaction: HRI’09*, 2009, pp. 53–60.

[18] C. L. Crown, “Coordinated interpersonal timing of vision and voice as a function of interpersonal attraction,” *Journal of Language and Social Psychology*, vol. 10, no. 1, pp. 29–46, 1991.

[19] B. Mutlu, J. Forlizzi, I. Nourbakhsh, and J. Hodgins, “The use of abstraction and motion in the design of social interfaces,” in *Proceedings of DIS’06*, ACM Press, 2006.

[20] G. Hoffman, “Evaluating fluency in human–robot collaboration,” *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 209–218, 2019.

[21] A. Bliek, S. Bensch, and T. Hellström, “How can a robot trigger human backchanneling?” In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, IEEE, 2020, pp. 96–103.

[22] S. Stedtler, V. Fantasia, T.-A. Tjöstheim, B. Johansson, I. Brinck, and C. Balkenius, “Is there really an effect of time delays on perceived fluency and social attributes between humans and social robots? a pilot study,” in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 1013–1017.

[23] S. Stedtler, *The impact of time delays on gaze behavior in human-robot interaction*, Retrieved January 6, 2025, from https://portal.research.lu.se/en/activities/the-impact-of-time-delays-on-gaze-behavior-in-human-robot-interac, 2024.

[24] I. Brinck and S. Stedtler, “Frictional design in human-robot interaction: Delayed movement in a turn-taking game,” in *Joint Proceedings of the Workshops at the Third International Conference on Hybrid Human-Artificial Intelligence: HHAI-WS 2024*, CEUR-WS, 2024.

[25] M. Robotics, *Misty ii*, Retrieved January 6, 2025, from https://mistyrobotics.com, 2024.

[26] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, “Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots,” *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.