Emotionally Specific Backchanneling in Social Human-Robot Interaction and Human-Human Interaction

Pourya Shahverdi[1], Katelyn Rousso[1], Justin Klotz[1], Iman Bakhoda [1], Malek Zribi[1], Wing-Yue Geoffrey Louie[1]

Abstract— **Backchanneling models, designed to enhance the interactive capabilities of robots, have primarily been trained on human-human interaction data. However, applying such data directly to social robots raises concerns due to dissimilarities in the way humans and robots exhibit verbal and nonverbal behaviors, particularly in the domain of emotional backchannels. This research aims to address this gap by conducting an exploratory study on the differences in human backchanneling behaviors during interactions with humans and social robots in various emotional contexts (e.g., happy and sad). Our findings reveal significant variations in emotionally specific backchannels between human-human and human-robot interactions under different emotional contexts. These results highlight the importance of designing backchanneling models that are tailored for human-robot interactions.**

I. Introduction

Backchanneling plays an essential role in a conversation and contributes to the naturalness of social Human-Robot Interaction (HRI). Backchanneling refers to the verbal phrases and non-verbal behaviors that listeners use to indicate that they are attending to a conversation, engaged, emotionally affected, and following along with the speaker. Backchanneling can include lexical phrases such as ”yes”, ”okay”, ”perfect”, and ”alright” as well as non-verbals such as nodding, eye gaze, or non-lexical sounds such as ”ha, ”oh”, ”mm-hmm”, ”uh-huh”. A combination of both verbals and non-verbals could also be used such as saying ”yes” while nodding [1], [2].

These backchannels can be categorized into two types: generic and specific. Generic backchannels (GBCs) refer to feedback that is not specific to the context and is considered as a standard response. Examples of generic backchannels include nodding or verbal phrases such as ”yeah” and could be classified as continuers since they help to maintain the flow of the conversation by signaling to the speaker that the listener is engaged and interested [3]. Specific backchannels (SBCs) refer to feedback that is tailored to the context and offers more information about the listener’s understanding of the situation and emotion during the conversation. For example, when people listen to a sad story, they often show sadness through their facial expressions and utter sounds such as ”oh” or ”ah”. Whereas, they may display excited facial expressions and sounds like ”yea” or ”ha-ha” to show their happiness and empathy when they hear good news. These are known as emotionally specific backchannels and are used to convey empathy and affective responses to a story [4].

Using the appropriate backchannel based on the emotional context of the conversation can make a speaker feel more comfortable, heard, and valued, which can lead to a more empathetic and engaging conversation overall. On the other hand, a lack of backchannels or inappropriate emotional backchannels can negatively influence the conversation flow and engagement because a speaker may feel disrespected, bored, frustrated, or annoyed [5], [6]. As robots become more integrated into various aspects of society, it’s also important to consider the ethical implications of their use. For example, if robots are used in healthcare settings, it’s important to ensure that they are able to engage in empathetic and supportive conversations with patients. Hence, it is necessary that social robots are too capable of producing appropriate backchanneling behaviors according to the emotional context of HRIs [7].

One challenge with designing appropriate backchanneling behaviors for human-robot interactions is that social robots often produce verbal and non-verbal behaviors in different ways then humans. For instance, some robots use LED colors to communicate emotions [8]. In some cases, social robots produce facial expressions by adjusting only their eyebrows and lip shape (e.g., iCat, Cozmo) [9], [10]. While some social robots are able to produce human-like facial expressions they utilize different mechanisms to achieve these expressions such as through projections or wire actuation of a rubber skin (e.g., Furhat, Sophia, and Ameca) [11]–[13]. Although social robots can produce emotionally-specific verbal and non-verbal backchannels to exhibit emotional responsiveness and engagement in human-robot interactions, they may not be perceived in the same manner as when they are produced in human-human interactions and this can potentially reduce the overall effectiveness of these human-robot social interactions [14].

Due to all the differences between social robots and humans in performing verbal and nonverbal behaviors to communicate, the way humans perceive, engage, and interact with social robots may differ from their interactions with humans. This can impact the performance of backchanneling models designed for human-robot interactions because these models are often trained on human-human interaction datasets [1], [15]–[18]. Namely, if a machine learning model doesn’t receive appropriate and authentic input behaviors from human users (i.e., states), it may struggle to generate accurate outputs

This work was supported by the National Science Foundation grants #1948224 and #2238088

[1]Intelligent Robotics Laboratory, Oakland University, Michigan, USA, 48309 (e-mail: louie@oakland.edu)

Digital Object Identifier (DOI): see top of this page.

(i.e., backchanneling behavior). Hence, shedding light on the differences between social robots and humans in triggering emotionally-specific backchanneling can lead to improved backchanneling models in the future.

This research aims to investigate how human backchanneling behavior differs when interacting with a robot compared to interacting with a human in different emotional contexts. In order to investigate these differences, an exploratory study was conducted that compares the generic and emotionally-specific backchanneling behaviors of humans while interacting with an interactive social robot or human storyteller in two different emotional contexts. The emotional contexts (happy and sad) were designed to elicit emotionally-specific backchanneling from the participants. The human-human and human-robot interactions were then analyzed to investigate the differences in participants' backchanneling behavior.

II. Related works

Social robotics researchers focus on training models on human backchanneling behavior in order to improve the naturalness of conversation in social HRIs. As a widespread approach, some studies used human-human interaction datasets to train a social robot to produce human-like backchanneling behaviors and evaluated the efficacy of their data-driven backchanneling behavior through a HRI user study. For example, [15] trained head-nodding generic backchanneling behavior using data augmentation, and [16] presented smiling and nodding backchanneling models that were trained using batch reinforcement learning. They compared a data-driven model against a rule-based model and a baseline model that produced random behaviors. The results of both studies indicated that participants showed a clear preference for the backchannels produced under the data-driven model when compared to the rule-based model or the baseline model. However, it's important to note that they did not explore how the effectiveness of these backchanneling behaviors might be influenced by the context of the interactions between humans and robots.

In [19], researchers explored how using three non-lexical backchannels in the Korean language (e.g., ne, eo, eum) by a Mebot social robot in two conversational contexts (e.g., information-centric and emotion-centric) are rated by 96 online observers. The results of this study showed that the conversational context of a non-lexical backchannel affects the rater's perception of a robot and is an important factor to consider when evaluating the effectiveness of backchanneling behaviors. Although this study compared the effects of using non-lexical backchannels in information-centric to emotion-centric conversational contexts, different emotional contexts were not considered. Moreover, this study had only one robot's interactant and did not target different robot users' backchanneling behavior to evaluate. People may not have the same level of emotional reactions, empathy, or self-disclosure while talking to a robot or human. Consequently, a human may backchannel differently when interacting with a robot rather than a human in different emotional contexts.

As seen in [15], [16], [19], HRI studies currently investigate how a robot's backchanneling behavior affects a speaker's engagement and perception towards a social robot. In contrast with these studies, the research in [20] investigated a listener's backchanneling behaviors while they were interacting with a Pepper social robot, which was capable of exhibiting backchanneling-inviting cues such as pausing, gazing, and gesturing. The results of this research found that a listener's backchanneling behavior differs significantly when interacting with a robot than a human. However, the robot was unable to adapt its backchannel feedback to the listener. Addressing such limitations is important because studies have observed that users are more engaged with a robot listener capable of producing adaptive backchannels that express a robot's attentiveness to the speaker [21]. Additionally, in [20], the social robot Pepper was limited to exhibiting backchanneling-inviting cues such as pausing, gazing, and gesturing and was not able to show human-like facial expressions, whereas more advanced social robots can show specific emotions and maintain human-like eye contact with a listener. For example, the Furhat robot can produce many different facial expressions depending on the emotional context, such as smiling, frowning, and raising eyebrows. Facial expressions of emotion can influence a human's perception and trustworthiness of a social robot, and therefore a human may feel more invited to produce backchannels and engage with a robot capable of human-like facial expressions. Thus, it may be useful to explore a human's backchanneling when interacting with a social robot capable of producing human-like facial expressions.

Current research in social robotics investigates both generic and specific backchanneling behaviors in order to produce a more engaging and natural human-robot interaction [15], [16], [19]–[21]. Many of these studies tend to investigate the efficacy of data-driven backchanneling behaviors on a human's engagement and perception of a social robot [15], [16]. To the best of our knowledge, there has been a lack of investigation on a human's emotionally-specific backchanneling while they interact with a social robot. Although [20] investigates a human's backchanneling behaviors when interacting with the Pepper robot, the use of a non-facial expressive robot in this study may influence a human's perception and trustworthiness of a social robot and production of backchannels. Additionally, humans may not feel the same level of empathy towards a robot as a human, and a human may backchannel differently when interacting with a robot than a human in different emotional contexts. The study presented in this paper aims to explore how people's backchanneling behavior differs when interacting with a human and a social robot that is capable of exhibiting natural human-like facial expressions.

III. Experimental Design

This study employed a 2x2 mixed design to investigate differences in listener generic and specific backchannels when interacting with a robot or a human storyteller narrating stories with different emotional contexts. The between-subjects condition in this design was the robot or human storyteller.

The within-subjects condition was the emotional context of the story being told which included happy or sad. Participants were randomly assigned to conditions.

A. Interactive Story Telling

An interactive storytelling scenario was designed as it provides opportunities for a participant to engage in the interaction as both a listener and speaker as well as elicits generic and specific backchannels from the participant. This is accomplished by allowing the audience to actively participate in a story and shape its flow through their choices. Furthermore, an emotionally expressive social robot can elicit different emotional responses in a listener during storytelling [22], [23].

Two stories, sad and happy, were developed and interactively presented by a robot and a human in this experiment. The distinction between happy and sad stories rested on the elements of the story. A happy story is typically one that evokes positive emotions such as joy, contentment, and satisfaction within the reader/listener. A sad story is one that evokes negative emotions such as grief, sadness, or disgust within the reader/listener. In both conditions, the participants were given options to choose the actions of the character in the story. Overall, the stories were 8-15 minutes in length depending on the how participants responded.

In the sad interactive story condition, a dog named Lacy was injured by a car, and the participant had to choose how the professor of veterinary science (the main character) would treat the dog. As the story progresses, the professor and Lacy build a close bond as a result of rehabilitating her. As a turn of events, despite the close relationship between the professor and Lacy, the professor is unable to keep Lacy at the end of the story.

The happy interactive story condition had the participant decide how a linguistics professor would navigate through teaching a language he did not know how to speak. The professor was constantly faced with situations where he had to keep lying about knowing the language. However, in the end, his poor situation became positive despite not fluently speaking the language that he taught.

To ensure participants' were able to distinguish sad from happy content and that their emotions reflected the content of the stories, literary techniques were used in the story formulation to help effectively elicit emotions at the appropriate times while engaging the readers. This included using emotional turning points as well as emotional climaxes in order to shift the tone in the story and attempt to induce specific emotional backchannels [24]. For example, a positive turning point in the happy story was when the student pointed out that the teacher was teaching the wrong language class. This was strategically placed in order to evoke a Positive Specific Backchannel (PSBC) to go along with the moment of a happy surprise. Conversely, an emotional turning point within the sad story was when Lacy was hit by a car for the second time in order to evoke a Negative Specific Backchannel (NSBC) to go along with the sad moment of shock. It should be noted that researchers sought to increase the effectiveness of the emotional turning points by having the remaining events in the story provide neutral emotional content. There were 6 emotional turning points in the sad story and 7 for the happy story.

TABLE I: Backchannels used by the storytellers

Emotion	Level	Backchannel
Happy	High	Smile, nod, and say "thanks"
Happy	Medium	Smile, nod, and say "okay"
Happy	Medium	Smile, nod, and say "nice"
Happy	Low	Smile and nod
Neutral	-	Nod and say "okay"
Neutral	-	Nod and say "umm"
Neutral	-	Nod
Neutral	-	Say "alright"
Sad	Low	Express sad and look down
Sad	Medium	Say "that's sad", shake head, and look down
Sad	Medium	Say "I can't believe that", shake head, and look down
Sad	High	Shake head, express sad, and say "oh no!"

The emotional climax strategy aimed to build up tension in the story and increase the probability for specific emotional backchannels. For example, In the happy story, the emotional climax was when the professor was invited to a dinner party with his colleagues who speak the language he doesn't know very well. The sad story's climax was when Lacy bit one of the other animals that the professor owned and was forced to leave her on the doorstep of an animal shelter. While the emotional climaxes are reaching their peak, it is also the point in the story where the reader should be the most engaged.

Two more techniques used to evoke specific emotional backchannels were varying the tone and specific emotional backchannels of the speaker. This meant that in parts of the story where an emotional turning point or climax occurred, the speaker's tone would change to make the situation more dramatic (e.g., a sad story with a sad event would have the speaker say that part in a sad or terrified tone). The speaker also used 4 happy, 4 sad, and 4 neutralbackchannels. The happy and sad backchannels had varying levels of intensity (i.e., 1 low level, 2 medium level, and 1 high level). Descriptions for the backchannels are presented in Table I. Lastly, in order to keep the reader engaged and equalize turn-taking during the interaction, open-ended questions were used during the interaction (e.g., "Why did you choose to do that?").

All in all, each story would be between 8-15 minutes in length depending on how much the participant was willing to speak. One thing that should be acknowledged is the reasoning for uneven ratios of the open-ended and emotional turning points. This was the case as our pilot experiments determined that for the specifically written stories, these ratios would be optimal for testing and achieving the desired results.

B. Measures and Hypotheses

To address the research question within this study, we defined the measures and metrics as described in Table II.

While exploratory research aims to collect evidence to form hypotheses, it can be further improved by initially proposing a set of working hypotheses beforehand, which can be later refined and modified based on the findings [25]. Our working hypotheses are presented as follows:

WH1: Participants will produce a significantly higher number of generic backchannels and specific backchannels

TABLE II: Measures and metrics

Backchanneling			
Metric	**Label**	**Definition or Example**	**Type**
Generic Backchannel	GBC	Backchannels with a neutral emotion (e.g., nodding or saying ”ok”)	Frequency
Specific Backchannel	SBC	Includes positive and negative emotional backchannels	Frequency
Positive Specific Backchannel	PSBC	SBCs that are used in positive emotions like happinnes	Frequency
Negative Specific Backchannel	NSBC	SBCs that are used in negative emotions like sadness	Frequency
Percentage of GBCs	%GBCs	GBCs / (GBCs + SBCs) x 100	%
Percentage of SBCs	%SBC	SBCs / (GBCs + SBCs) x 100	%
Percentage of PSBCs	%PSBC	PSBCs / (PSBCs + NSBCs) x 100	%
Percentage of NSBCs	%NSBC	NSBCs / (PSBCs + NSBCs) x 100	%
Participant’s Evaluation of the Story’s Emotion			
Evaluation of the Happy Story	EoH	Participants rate the happy story from 1 to 9	Likert Scale
Eavaluation of the Sad Story	EoS	Participants rate the sad story from 1 to 9	Likert Scale

during interactions with humans compared to the robot storytellers’ condition.

WH2: Participants’ evaluations of the emotions for the happy story will be significantly higher than their evaluations of the sad story.

WH3: The number of participants’ positive specific backchannels and their percentage in the happy condition will be significantly higher than the positive specific backchannels and their percentage in the sad condition.

WH4: Participants’ negative specific backchannels and their percentage in the happy condition will be significantly lower than the negative specific backchannels and their percentage in the sad condition.

WH5: Participants’ negative specific backchannels and their percentage in the sad condition will be significantly higher than the negative specific backchannels and their percentage in the happy condition.

WH6: Participants’ positive specific backchannels and their percentage in the sad condition will be significantly lower than the positive specific backchannels and their percentage in the happy condition.

WH7: There is a significant difference in the number of each specific backchannel (e.g., positive, negative) between the robot and the human storytellers for both emotional contexts.

C. Experimental Setup

In the human storyteller condition, the narrator was one of the researchers. The researcher was a 20-year-old American female that is a native English speaker. Figure 1 shows the human storyteller on the right and the participant on the left. For the robot storyteller condition, a Furhat social robot (Figure 2) was used and teleoperated by a researcher utilizing a customized Wizard of Oz (WoZ) interface to narrate the story. The teleoperator was able to observe the interaction through a one-way mirror in a neighboring room (Figure 2).

1) Furhat Social Robot: The Furhat robot consists of an animated human-like face projected on a physical mask on the back of its head [11]. This approach is referred to as ”blended embodiment” [26] and produces more realistic facial expressions, eye gaze, and human-like appearances. The robot also has the ability to move its head in a natural way, similar to a human, through its human-like neck design.

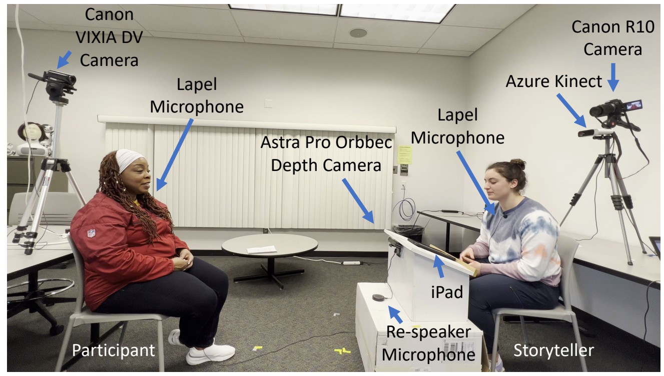

Fig. 1: Experimental Setup for the Human Storyteller Condition

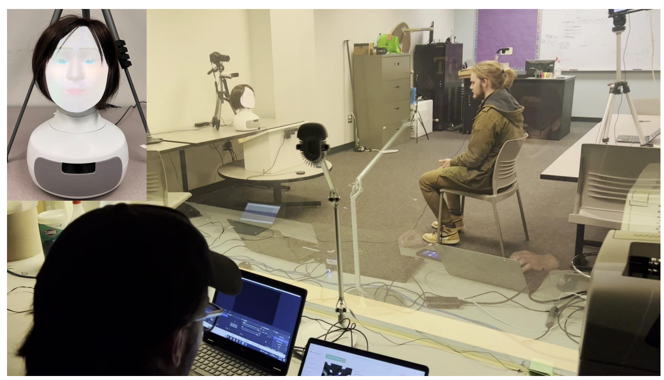

Fig. 2: WoZ Setup and the Furhat Social Robot

It is also equipped with a 135-degree field-of-view camera, 2 microphones, and a speaker. The robot can also switch between different projected faces and physical masks. For the voices, a wide selection of options is available such as Amazon Polly and Acapela Group. According to the ABOT database, which measures human-like appearance, Furhat has a score of 63.43, which includes scores for body manipulators (0.025), facial features (1.0), and surface (0.727). This score is higher than those of other similar robots like NAO (45.92) and Pepper (42.17), even though Furhat does not have arms and legs [27].

D. Procedure

The experiment procedures in this study were approved by the Institutional Review Board (IRB) at Oakland University (#IRB-FY2022-103). A total of 28 participants were recruited and 14 participants were assigned to each storyteller condition. There were 15 males and 13 females with an average age of 30.

Prior to beginning the experiment, a researcher explained all the procedures to participants in a separate room from the storyteller and written informed consent was obtained. The participant was then asked to fill out a questionnaire regarding their demographic information, and their experience with technology. The participant was then guided to the room and seated in front of the storyteller. The first story was then narrated by the storyteller. After the first story, the participant was asked to fill out a questionnaire on rating the story's emotion on a Likert scale from 1-9 where 1 was very sad and 9 was very happy. The second story was then narrated by the storyteller and participants were asked to rate the second story as well. During the story narrations, the story was printed and placed on a desk in front of the storytellers. Since the human storyteller was unable to memorize the entire interactive story, she was allowed to use the manuscript on an iPad to conduct the narration. To keep both storyteller conditions consistent, the robot was programmed to pretend to also read through the story on paper at the same points as the human storyteller.

E. Data Collection

The interactive story interactions were all recorded. All devices used for data collection are depicted in Figure 2. In both conditions, an Orbbec Astra [1] and a Microsoft Azure Kinect[2] were used for collecting depth data from two different perspectives during the interaction (shown in Figure 1). A video camera (Canon R10) was also recording a closeup view of the participant and another camera (Canon VIXIA) was recording the storyteller. Two lapel microphones were used to record the vocal data from the participants and the storytellers. Moreover, audio data was recorded using a Respeaker microphone array [3].

F. Data-Coding

The videos from the participants were used to code their backchanneling behavior. Their backchanneling behavior was coded by researchers based on the labels defined in Table II. The session was divided into ten-second intervals and coded for instances of GBCs, PSBCs, and NSBCs. Since all labels could be displayed within a ten-second interval, it was possible for an interval to be categorized as GBC, PSBC, or NSPC.

Inter-Rater Reliability (IRR): 50% of the interaction sessions were double-coded and rated by two researchers (i.e., raters) to ensure that the IRR was over 70%. IRR was scored using the interval-by-interval method for all the labels. This method measures the percentage of matched intervals with a label. IRR scores below 70% were corrected by the two raters coming together to re-code the session [28].

[1]https://orbbec3d.com

[2]https://azure.microsoft.com/en-us/products/kinect-dk

[3]Datasets will be available online upon request from the corresponding author

TABLE III: Results of the Statistical Analysis

		μ	σ	t	p	Result
WH1	GBCs in Robot	7.32	7.10	1.43	0.078	Rejected
	GBCs in Human	4.96	5.01			
WH1	SBCs in Robot	11.92	8.67	3.77	<0.001	Accepted
	SBCs in Human	5	4.3			
WH2	EoH	6.32	1.61	10.29	<0.001	Accepted
	EoS	2.10	1.59			
WH3	PSBCs in Happy	8.53	6.20	3.67	<0.001	Accepted
	PSBCs in Sad	4.28	6.99			
WH3	%PSBCs in Happy	0.86	0.25	3.94	<0.001	Accepted
	%PSBCs in Sad	0.52	0.36			
WH4	NSBCs in Happy	1.85	0.35	-3.8	<0.001	Accepted
	NSBCs in Sad	3.77	0.71			
WH4	%NSBCs in Happy	0.09	0.19	-4.9	<0.001	Accepted
	%NSBCs in Sad	0.44	0.3			
WH5	NSBCs in Sad	3.14	3.77	3.80	<0.001	Accepted
	NSBCs in Happy	0.96	1.85			
WH5	%NSBCs in Sad	0.44	0.36	4.91	<0.001	Accepted
	%NSBCs in Happy	0.09	0.19			
WH6	PSBCs in Sad	4.28	6.99	-3.67	<0.001	Accepted
	PSBCs in Happy	8.53	6.20			
WH6	%PSBCs in Sad	0.52	0.36	-3.94	<0.001	Accepted
	%PSBCs in Happy	0.86	0.25			
WH7	PSBCs in Robot	8.92	7.88	2.91	0.003	Accepted
	PSBCs in Human	3.89	4.63			
WH7	NSBCs in Robot	3	3.9	2.34	0.011	Accepted
	NSBCs in Human	1.10	1.7			

IV. Data Analysis

We used paired sample t-tests to evaluate WH2, WH3, WH4, WH5, and WH6. Independent sample t-tests were used to evaluate WH1 and WH7. The independent variables for the t-tests were the stories' emotions (happy and sad), and the storytellers' (robot and human). The dependent variables for our statistical analysis are listed in Table II.

V. Results and Discussion

The findings of our study, as summarized in Table III, have provided valuable insights into the dynamics of human-robot interactions and their impact on backchanneling behaviors. One of the primary hypotheses explored in this research was whether there existed a significant difference in participants' GBCs (Generic Backchanneling Cues) between interactions with humans and robots, irrespective of the emotional contexts. Contrary to our initial expectations, the results led us to reject this hypothesis, suggesting that the storytelling scenario was well-balanced in terms of generic content delivery, resulting in similar GBCs from participants across both human and robot interactions.

However, the second hypothesis, which examined participants' evaluations of the stories, was supported by the data. Notably, participants perceived the emotional tone of the stories differently, with the happy story evoking a genuinely happy response and the sad story eliciting a sense of sadness. This finding demonstrates that emotional storytelling can indeed influence participants' emotional engagement, even in interactions with social robots.

Moving on to hypotheses 3, 4, 5, and 6, we aimed to objectively evaluate the stories' ability to induce emotionally specific backchanneling cues in participants. The results of the statistical analysis revealed a significant difference in participants' specific backchannels between the happy and sad emotional contexts. This highlights the importance of considering emotional factors when designing and delivering stories in human-robot interactions, as these cues play a crucial role in shaping users' emotional responses.

With evidence now demonstrating both subjective and objective differences in emotional contexts, hypothesis 7, targeting the main research question of this paper, was supported. The statistical analysis revealed that participants' emotionally-specific backchannels were significantly different across the human and robot interactions. This suggests that social robots may indeed elicit distinct emotional responses from users compared to human counterparts, and understanding these differences is vital for developing more effective and emotionally resonant human-robot interactions.

Given the implications of our results, it becomes evident that training machine learning models for predicting and generating backchanneling behaviors, particularly those that are emotionally specific, based solely on human-human interaction data may not be the optimal approach. Instead, incorporating data from human-robot interactions may be crucial to developing more contextually aware and emotionally sensitive backchanneling models for social robots.

As we look to the future, our research will extend into training backchanneling models specifically on human-robot interaction data. By comparing the performance of these models against the baseline model trained on human-human interaction data, we aim to gain further insights into the effectiveness and appropriateness of using human-robot interaction data for such models. This endeavor will contribute to refining the design and implementation of social robots, ultimately enhancing their ability to engage users emotionally and foster more natural and compelling interactions.

REFERENCES

[1] A. I. Adiba, T. Homma, and T. Miyoshi, "Towards immediate backchannel generation using attention-based early prediction model," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 7408–7412.

[2] N. Ward and W. Tsukahara, "Prosodic features which cue back-channel responses in english and japanese," *Journal of pragmatics*, vol. 32, no. 8, pp. 1177–1207, 2000.

[3] K. M. Senk, "Analyzing interactional management in native/non-native english conversation: A case of listener response," *IRAL: International Review of Applied Linguistics in Language Teaching*, vol. 35, no. 1, p. 37, 1997.

[4] B.-J. Theobald, I. Matthews, M. Mangini, J. R. Spies, T. R. Brick, J. F. Cohn, and S. M. Boker, "Mapping and manipulating facial expression," *Language and speech*, vol. 52, no. 2-3, pp. 369–386, 2009.

[5] J. B. Bavelas, L. Coates, and T. Johnson, "Listeners as co-narrators." *Journal of personality and social psychology*, vol. 79, no. 6, p. 941, 2000.

[6] J. Tolins and J. E. F. Tree, "Addressee backchannels steer narrative development," *Journal of Pragmatics*, vol. 70, pp. 152–164, 2014.

[7] B. Lugrin, C. Pelachaud, and D. Traum, *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics, Volume 2: Interactivity, Platforms, Application*. Morgan & Claypool, 2022.

[8] D. O. Johnson, R. H. Cuijpers, and D. van der Pol, "Imitating human emotions with artificial facial expressions," *International Journal of Social Robotics*, vol. 5, no. 4, pp. 503–513, 2013.

[9] A. van Breemen, X. Yan, and B. Meerbeek, "icat: an animated user-interface robot with personality," in *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, 2005, pp. 143–144.

[10] H. R. Pelikan, M. Broth, and L. Keevallik, "" are you sad, cozmo?" how humans make sense of a home robot's emotion displays," in *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2020, pp. 461–470.

[11] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granström, "Furhat: a back-projected human-like robot head for multiparty human-machine interaction," in *Cognitive behavioural systems*. Springer, 2012, pp. 114–130.

[12] J. Retto, "Sophia, first citizen robot of the world," *ResearchGate, URL: https://www. researchgate. net*, 2017.

[13] Ameca. [Online]. Available: https://www.engineeredarts.co.uk/robot/ameca/

[14] D. Lala, K. Inoue, P. Milhorat, and T. Kawahara, "Detection of social signals for recognizing engagement in human-robot interaction," *arXiv preprint arXiv:1709.10257*, 2017.

[15] M. Murray, N. Walker, A. Nanavati, P. Alves-Oliveira, N. Filippov, A. Sauppe, B. Mutlu, and M. Cakmak, "Learning backchanneling behaviors for a social robot via data augmentation from human-human conversations," in *Conference on Robot Learning*. PMLR, 2022, pp. 513–525.

[16] N. Hussain, E. Erzin, T. M. Sezgin, and Y. Yemez, "Training socially engaging robots: Modeling backchannel behaviors with batch reinforcement learning," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 1840–1853, 2022.

[17] R. Ruede, M. Müller, S. Stüker, and A. Waibel, "Yeah, right, uh-huh: a deep learning backchannel predictor," in *Advanced Social Interaction with Agents*. Springer, 2019, pp. 247–258.

[18] N. Hussain, E. Erzin, T. M. Sezgin, and Y. Yemez, "Speech driven backchannel generation using deep q-network for enhancing engagement in human-robot interaction," *arXiv preprint arXiv:1908.01618*, 2019.

[19] S. Kim, S. Seok, J. Choi, Y. Lim, and S. S. Kwak, "Effects of conversational contexts and forms of non-lexical backchannel on user perception of robots," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 3042–3047.

[20] A. Bliek, "Backchanneling in human-robot interaction," Ph.D. dissertation, University of Groningen, 2020.

[21] 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.

[22] S. Costa, A. Brunete, B.-C. Bae, and N. Mavridis, "Emotional storytelling using virtual and robotic agents," *International Journal of Humanoid Robotics*, vol. 15, no. 03, p. 1850006, 2018.

[23] M. Spitale, S. Okamoto, M. Gupta, H. Xi, and M. J. Matarić, "Socially assistive robots as storytellers that elicit empathy," *ACM Transactions on Human-Robot Interaction*, 2022.

[24] D. Maass, *The emotional craft of fiction: How to write the story beneath the surface*. Penguin, 2016.

[25] M. Casula, N. Rangarajan, and P. Shields, "The potential of working hypotheses for deductive exploratory research," *Quality & Quantity*, vol. 55, no. 5, pp. 1703–1725, 2021.

[26] M. Paetzel and G. Castellano, "Let me get to know you better: can interactions help to overcome uncanny feelings?" in *Proceedings of the 7th International Conference on Human-agent Interaction*, 2019, pp. 59–67.

[27] E. Phillips, X. Zhao, D. Ullman, and B. F. Malle, "What is human-like? decomposing robots' human-like appearance using the anthropomorphic robot (abot) database," in *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, 2018, pp. 105–113.

[28] N. Gisev, J. S. Bell, and T. F. Chen, "Interrater agreement and interrater reliability: key concepts, approaches, and applications," *Research in Social and Administrative Pharmacy*, vol. 9, no. 3, pp. 330–338, 2013.

,