Toward Designing a Toolkit for Intuitive Design of Backchanneling Behaviour in Social Robots

Samuel R. M. Barrett
Department of Computer Science
University of Manitoba
Winnipeg, Canada
barret15@myumanitoba.ca

James E. Young
Department of Computer Science
University of Manitoba
Winnipeg, Canada
young@cs.umanitoba.ca

***Abstract*—Conversational social robots can use backchanneling to provide real-time feedback, convey an understanding, demonstrate attention and maintain a flowing dialog. However, backchanneling implementations are often project- and robot-specific, and we do not yet have standardized backchanneling toolkits to enable robot dialog designers to easily configure a robot's backchannel behaviour. This highlights a need for high-level toolkits to enable dialog designers to engage with backchanneling and customization, allowing rapid exploration of backchanneling as a part of dialog design. To engage with this problem, we surveyed recent works on backchanneling social robots, performed an initial analysis to describe the range of backchannel techniques, and outlined preliminary criteria for high-level backchanneling design.**

Keywords— Backchanneling; Social Robots.

I. Introduction

Backchanneling is a listening behaviour used in conversations between people to convey feedback to a speaker, for instance with a verbal utterance such as "yes" or, "hmm," or with non-verbal actions like nodding one's head [1]. This feedback plays an important role in maintaining the flow of dialog and enables the listener to communicate attentiveness, understanding, and agreement [2], [3], [4]. In this way, backchanneling keeps each party informed of the current state of the interaction, especially for turn-taking, the collaborative process that regulates which party is to speak [5]. Given the utility of backchanneling in human-human conversations, roboticists have been employing these feedback behaviours in robots for improved human-robot dialog.

Recently, backchanneling has become a focus of dialog system implementations. Verbal utterances have commonly been used in conversational agents, where speech is a prominent communication modality. Lala et al. [6] demonstrated an agent that backchannels as a listening third party in a conversation, contributing semantic utterances and laughter. Blomsma et al. [7] assessed participant perceptions of a virtual robot's personality for different backchanneling behaviours. Compared to virtual agents, social robots feature a broader range of communication modalities and their physical embodiment elicits stronger reactions from people [8], lending unique considerations to backchanneling in robots.

A social robot's embodied nature affords a variety of communication modalities, encouraging design of multimodal backchanneling behaviours. In one project, a Nao robot behaved as an active listener by gazing at the speaker and nodding or shaking its head [9]. Tatarian et al. [10] implemented a multimodal conversational robot which used gaze aversion cues in addition to nodding and verbal backchannels. Park et al. [11] investigated the use of smiles, head nods and eye movements in a listening robot for children's storytelling. The effects of head-nodding backchannels were examined for a virtual agent and a social robot [12]. Kawahara developed multimodal backchanneling behaviours for attentive listening in the robot system ERICA [13]. Roboticists have used a range of modalities and techniques for backchanneling in social robots which we must consider in the development of a dialog-creation toolkit.

Robot designers leverage backchanneling techniques toward achieving desired interaction goals. Backchanneling was used to encourage a person to speak up or to continue speaking, for example when dialog participation was imbalanced in a group scenario [12], [14]. Social robots were designed for attentive listening behaviours, leveraging backchannels to communicate attention and understanding toward a speaker [6], [11]. Further, backchanneling helps social robots achieve natural dialog and fluent turn-taking in a conversation [1], [13], [15]. Smith et al. [2] proposed backchanneling to reduce cognitive load on users in conversations with artificial agents. Backchanneling has also been shown to impact a user's perceived trust in a robot [9]. Backchanneling is employed in social robots for myriad purposes using multiple modalities, indicating a need for flexibility of backchanneling design in a toolkit for creating robot backchanneling behaviour.

Given the plethora of backchanneling robot implementations, we seek to create a structured approach to framing and developing backchannel behaviours. In this work, we begin to engage this problem by first conducting an initial literature survey to identify research projects implementing backchanneling in robots. We then analyzed each paper to extract key themes from implementations, resulting in an initial framework for describing the range of backchanneling behaviours utilized by robot designers. We will expand on the details of these backchanneling behaviours in a future work to inform our development of a toolkit for creating customized backchanneling in social robot dialog.

II. Methodology

We surveyed two key publication venues in human-robot interaction, the ACM/IEEE Conference on Human-Robot Interaction, and the ACM Transactions on Human-Robot Interaction Journal. We conducted a keyword search using the query "backchannel" on the ACM Digital Library to identify relevant papers. Following, we filtered these papers to remove duplicates and to remove works that only mentioned backchanneling in passing. Finally, we analyzed the backchanneling behaviours employed by the robot from the perspective of dialog design, noting techniques of the backchanneling. We collected these into themes to form an initial framework for describing the range of backchanneling behaviours in social robots. This work will inform our approach to a toolkit for creating customized backchanneling behaviours in a social robot.

III. Preliminary Findings

Our search initially resulted in 58 papers. Filtering for duplicate projects and work without core backchanneling resulted in 18 papers. We examined these 18 papers to identify the range of backchanneling techniques used in human-robot interaction and organized them into natural categories of modalities (Table 1). We discuss these categories below.

Following, we organized the features from Table 1 into four main categories; verbal utterances, gaze, head movements, and multimodal backchannels.

Facial expressions were relatively uncommon in the literature and highly depend on the robot's physical capabilities. For example, Lee et al. [16] emulated eyebrow raises by widening a robot's eyes, which were displayed on a digital screen on its face. Additionally, while several implementations incorporated some form of body or arm gesture, these were mostly a collection of robot-specific behaviours and thus were not identified as recurring themes. We proceed by examining our extracted themes.

A. Verbal Utterances

We found that fewer than half of backchanneling implementations employed some kind of verbal utterance to communicate with a user. One of the challenges for verbal backchanneling in robots is to insert speech that does not feel like an interruption to the user. In one work, Chao and Thomaz expressed that supportive backchannels–meant to convey attentive listening and encourage participant speech–were perceived as the robot speaking over participants [17]. Similarly, Park et al. [11] observed that children were distracted from storytelling when a robot backchanneled randomly, an utterance such as "wheeee", instead of determining an appropriate time to speak. These results reflect the importance of timing in conversations, including for backchannel behaviour [11], [17], and suggest careful attention to timing is warranted.

Most implementations of verbal backchanneling used short, non-lexical utterances, such as "hmm", "uh-huh", and in one case, "wheeee," [18], [11], [16], [17], [19], [20]. These kinds of verbal backchanneling were used to convey attentive listening and understanding [11], [21], [22]. Additionally, they served as 'continuers' to encourage more speech from people [17], [22]. In these contexts, non-lexical utterances could be sufficient to achieve the desired effects of backchanneling. In other cases, robots used more advanced verbal backchanneling, for instance one implementation by Strohkorb Sebo et al. [20] which offered a supportive response contingent on specific keywords referenced by the user, such as "Screwdriver, okay", or otherwise fell back on a simpler utterance like "Uh huh," [20].

TABLE I. Prevalant Backchanneling Features

Backchanneling Feature	***Occurrences***
Verbal Utterances	7
Gaze Towards Person	13
Gaze Aversion	12
Nodding	12
Facial Expressions	4
Body or Arm Gestures	8

Within the surveyed implementations, the high prevalence of non-lexical utterances could be due to their broad applicability and ease of implementation. Tanaka et al. [23] noted how such backchannels required only pause detection to achieve a usable result, a simple "hai" along with a nod.

B. Gaze

Collectively, non-verbal backchannels were the most prevalent across designs, with gaze being the most commonly employed modality. One work found that a robot gazing at a human speaker contributed to perceived responsiveness, compared to avoiding the speaker's gaze by looking away [24]. Similarly, gazing at the speaker was used in multiple works to demonstrate attentive listening and encourage further speech [11], [21], [25], [26]. Attentive listening and responsiveness were the most frequent goals of robot gaze in an interaction and will be our focus for gaze in our work towards a backchanneling robot toolkit.

Gaze aversion is used in tandem with gaze to communicate with a dialog partner. Huang and Mutlu [27] created a backchanneling behaviour where a robot shifted its gaze from a person to an object that was referred to in speech. Another robot implementation used gaze aversion to communicate that it was under a cognitive load, making it look like it was busy thinking [28]. The robot thus appeared more thoughtful and deliberate to participants by averting its gaze at the start of a response to a question. One work implemented gaze aversion for multi-participant interactions and found that participants who were ignored by the robot's gaze spoke far less frequently than those who were targeted with gaze [29]. Going even further, Hoffman et al. [24] used gaze aversion in a social robot to create a negative response where the robot dismissed a participant's speech. We note that robot designers intentionally incorporate both directed gaze and gaze aversion in backchanneling behaviour to achieve different outcomes.

We observed that gaze was used to achieve diverse effects in an interaction. One work evaluated using human-like gaze behaviours in a storytelling robot led to better recall of information by participants, compared to non-human-like behaviour [30]. Another robot conveyed referent information using gaze by looking toward objects on a table when referenced by the robot in conversation [30]. Further, Utami and Bickmore [25] investigated multiparty interaction and used gaze directed to a pair of speakers who completed each other's turns. We identify gaze as a key technique for its prevalence in the surveyed works and for its applicability to a gamut of interaction goals.

C. Head Movements

Nodding was the most utilized head movement behaviour, often invoked in combination with another modality. Nodding backchannels were mainly used to express attentive listening and understanding. Chao and Thomaz [17] integrated nodding backchannels into a floor regulation model which encouraged the user to continue holding the speaking floor. Head nods can convey agreement in and acknowledgement [22], [24], [31]. Ishi et al. [32] reported that robot head nodding behaviours modelled after human nodding was perceived as more natural to participants. We add nodding to our list of techniques for a backchanneling social robot implementation.

By contrast, head shaking was uncommon, seen in only one surveyed backchanneling implementation [17]. Likewise, creating human-like head movements during idle behaviour was studied in just a couple of works [32], [33]. Regarding head movements, we will focus primarily on nodding as a core backchanneling technique, though we see fit to also include head shaking behaviours for flexibility in interaction design.

D. Multimodal Backchanneling

A key theme we extracted is the prevalence of multimodal backchanneling, which were employed in 14 robot backchanneling implementations. For instance, multiple works used verbal backchanneling in conjunction with head nods [20], [23], [24]. Many implementations using gaze did so as one aspect of a backchannel, holding gaze while delivering a verbal utterance or nod, see [11], [17], [24]. Backchanneling in this human-like, multimodal fashion is made possible by social robots' access to several human-like modalities. We identify this theme to inform our future work on a toolkit which will enable designing multimodal backchanneling.

We also noted that less common behaviours such as facial expressions, arm movements or pointing were typically reserved for use in a multimodal backchannel with at least one prevalent backchannel–verbal, gaze, or nodding. Mortimer, a robot for collaborative music-playing, exhibited facial expressions in conjunction with head motions [31]. Therefore, we note that certain robot-specific or otherwise less prevalent features like facial expressions or pointing can be implemented in a backchannel as a complement to one or more of the typical modalities.

IV. Conclusion and Future Work

In this work, we surveyed works on social robot implementations to identify the techniques used to achieve backchanneling. We extracted a set of key techniques for backchanneling, including verbal utterances, gaze, nodding, and the use of multimodal behaviours. This work contributes to informing our toolkit for designing backchannels in a social robot.

A. Backchanneling Toolkit

Our survey results highlight key features which we should consider building into a dialog-creation toolkit for social robots. We aim to expose these backchanneling techniques to dialog designers, enabling them to easily incorporate customized backchanneling behaviour in their dialog designs. First, while verbal backchanneling was less prevalent than gestures in the surveyed works, their significance for dialog and use in virtual agents prompt us to minimally include non-lexical verbal backchannels in our implementation criteria. We will investigate whether more complex utterances enabled by language understanding is necessary for the kinds of behaviours we are targeting. Further, we will explore how systems generate verbal backchannel utterances and note that timing is a key consideration for the effectiveness of backchanneling, as inappropriately timed backchannels could create interruptions or awkward delays.

We will incorporate nodding and gaze mechanisms to enable dialog designers to manage the interaction using gaze cues and head movements. These techniques will be available in tandem with verbal utterances to allow designers to leverage multimodal backchanneling in their designs. Our next major step is to build a toolkit that exposes these techniques in a social robot using a simple graphical user interface.

Acknowledgment

We acknowledge the Natural Sciences and Engineering Research Council of Canada, who funded this research through their Discovery Grants program, and the University of Manitoba for their funding through the Computer Science Entrance Award.

References

[1] G. Skantze, "Turn-taking in Conversational Systems and Human-Robot Interaction: A Review," *Comput. Speech Lang.*, vol. 67, p. 101178, May 2021, doi: 10.1016/j.csl.2020.101178.

[2] J. Smith *et al.*, "Towards Conversationally Intelligent Dialog Systems," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, New Orleans LA USA: ACM, Apr. 2022, pp. 1–7. doi: 10.1145/3491101.3519842.

[3] A. Axelsson, H. Buschmeier, and G. Skantze, "Modeling Feedback in Interaction With Conversational Agents—A Review," *Front. Comput. Sci.*, vol. 4, Mar. 2022, doi: 10.3389/fcomp.2022.744574.

[4] G. Wilcock and K. Jokinen, "Towards Increasing Naturalness and Flexibility in Human-Robot Dialogue Systems," in *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction: 10th International Workshop on Spoken Dialogue Systems*, E. Marchi, S. M. Siniscalchi, S. Cumani, V. M. Salerno, and H. Li, Eds., Singapore: Springer, 2021, pp. 109–114. doi: 10.1007/978-981-15-9323-9_9.

[5] H. Sacks, E. Schegloff, and G. Jefferson, "A Simplest Systematics for the Organization of Turn-Taking for Conversation," *Language*, vol. 50, no. 4, pp. 696–735, Dec. 1974.

[6] D. Lala, K. Inoue, T. Kawahara, and K. Sawada, "Backchannel Generation Model for a Third Party Listener Agent," in *Proceedings of the 10th International Conference on Human-Agent Interaction*, in HAI '22. New York, NY, USA: Association for Computing Machinery, Dec. 2022, pp. 114–122. doi: 10.1145/3527188.3561926.

[7] P. Blomsma, G. Skantze, and M. Swerts, "Backchannel Behavior Influences the Perceived Personality of Human and Artificial Communication Partners," *Front. Artif. Intell.*, vol. 5, Mar. 2022, doi: 10.3389/frai.2022.835298.

[8] S. H. Seo, D. Geiskkovitch, M. Nakane, C. King, and J. E. Young, "Poor Thing! Would You Feel Sorry for a Simulated Robot?: A comparison of empathy toward a physical and a simulated robot," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, Portland Oregon USA: ACM, Mar. 2015, pp. 125–132. doi: 10.1145/2696454.2696471.

[9] N. Anzabi and H. Umemuro, "Effect of Different Listening Behaviors of Social Robots on Perceived Trust in Human-robot Interactions," *Int. J. Soc. Robot.*, vol. 15, no. 6, pp. 931–951, Jun. 2023, doi: 10.1007/s12369-023-01008-x.

[10] K. Tatarian, R. Stower, D. Rudaz, M. Chamoux, A. Kappas, and M. Chetouani, "How does Modality Matter? Investigating the Synthesis

and Effects of Multi-modal Robot Behavior on Social Intelligence," *Int. J. Soc. Robot.*, vol. 14, no. 4, pp. 893–911, Jun. 2022, doi: 10.1007/s12369-021-00839-w.

[11] H. W. Park, M. Gelsomini, J. J. Lee, and C. Breazeal, "Telling Stories to Robots: The Effect of Backchanneling on a Child's Storytelling," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '17. New York, NY, USA: Association for Computing Machinery, Mar. 2017, pp. 100–108. doi: 10.1145/2909824.3020245.

[12] A. Krogsager, N. Segato, and M. Rehm, "Backchannel Head Nods in Danish First Meeting Encounters with a Humanoid Robot: The Role of Physical Embodiment," in *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*, M. Kurosu, Ed., Cham: Springer International Publishing, 2014, pp. 651–662. doi: 10.1007/978-3-319-07230-2_62.

[13] T. Kawahara, "Spoken Dialogue System for a Human-like Conversational Robot ERICA," in *9th International Workshop on Spoken Dialogue System Technology*, L. F. D'Haro, R. E. Banchs, and H. Li, Eds., Singapore: Springer, 2019, pp. 65–75. doi: 10.1007/978-981-13-9443-0_6.

[14] D. A. Kazzi and Vi. Winberg, *Adapting the backchanneling behaviour of a social robot to increase user engagement : A study using social robots with contingent backchanneling behaviour*. 2022. Accessed: Nov. 04, 2024. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-319361

[15] F. Gervits, R. Thielstrom, A. Roque, and M. Scheutz, "It's About Time: Turn-Entry Timing For Situated Human-Robot Dialogue," in *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, O. Pietquin, S. Muresan, V. Chen, C. Kennington, D. Vandyke, N. Dethlefs, K. Inoue, E. Ekstedt, and S. Ultes, Eds., 1st virtual meeting: Association for Computational Linguistics, Jul. 2020, pp. 86–96. doi: 10.18653/v1/2020.sigdial-1.12.

[16] J. J. Lee, F. Sha, and C. Breazeal, "A Bayesian theory of mind approach to nonverbal communication," in *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '19. Daegu, Republic of Korea: IEEE Press, Jan. 2020, pp. 487–496.

[17] C. Chao and A. L. Thomaz, "Controlling social dynamics with a parametrized model of floor regulation," *J Hum-Robot Interact*, vol. 2, no. 1, pp. 4–29, Feb. 2013, doi: 10.5898/JHRI.2.1.Chao.

[18] M. Gelsomini, H. W. Park, J. J. Lee, and C. Breazeal, "Engaging Children as a Storyteller: Backchanneling Models for Social Robots," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 407. doi: 10.1145/3029798.3036651.

[19] G. Angelopoulos and V. Estivill-Castro, "Human-Robot Dialogue that Elicits The Alignment of Moral Principles For Driverless Vehicles," in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 196–200. doi: 10.1145/3610978.3640641.

[20] S. Strohkorb Sebo, L. L. Dong, N. Chang, and B. Scassellati, "Strategies for the Inclusion of Human Members within Human-Robot Teams," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 309–317. doi: 10.1145/3319502.3374808.

[21] G. Angelopoulos and V. Estivill-Castro, "Human-Robot Dialogue that Elicits The Alignment of Moral Principles For Driverless Vehicles," in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 196–200. doi: 10.1145/3610978.3640641.

[22] Y. Sakai, Y. Nonaka, K. Yasuda, and Y. I. Nakano, "Listener agent for elderly people with dementia," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '12. New York, NY, USA: Association for Computing Machinery, 2012, pp. 199–200. doi: 10.1145/2157689.2157754.

[23] K. Tanaka, N. Yamashita, H. Nakanishi, and H. Ishiguro, "Teleoperated or Autonomous? How to Produce a Robot Operator's Pseudo Presence in HRI," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, in HRI '16. Christchurch, New Zealand: IEEE Press, Mar. 2016, pp. 133–140.

[24] G. Hoffman, G. E. Birnbaum, K. Vanunu, O. Sass, and H. T. Reis, "Robot responsiveness to human disclosure affects social impression and appeal," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '14. New York, NY, USA: Association for Computing Machinery, 2014, pp. 1–8. doi: 10.1145/2559636.2559660.

[25] D. Utami and T. Bickmore, "Collaborative user responses in multiparty interaction with a couples counselor robot," in *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '19. Daegu, Republic of Korea: IEEE Press, 2020, pp. 294–303.

[26] A. Ramachandran, C.-M. Huang, E. Gartland, and B. Scassellati, "Thinking Aloud with a Tutoring Robot to Enhance Learning," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 59–68. doi: 10.1145/3171221.3171250.

[27] C.-M. Huang and B. Mutlu, "The Repertoire of Robot Behavior: Designing Social Behaviors to Support Human-Robot Joint Activity," *J. Hum.-Robot Interact.*, vol. 2, no. 2, pp. 80–102, Jun. 2013, doi: 10.5898/JHRI.2.2.Huang.

[28] S. Andrist, X. Z. Tan, M. Gleicher, and B. Mutlu, "Conversational gaze aversion for humanlike robots," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, in HRI '14. New York, NY, USA: Association for Computing Machinery, Mar. 2014, pp. 25–32. doi: 10.1145/2559636.2559666.

[29] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita, "Footing in human-robot conversations: how robots might shape participant roles using gaze cues," in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, in HRI '09. New York, NY, USA: Association for Computing Machinery, 2009, pp. 61–68. doi: 10.1145/1514095.1514109.

[30] C.-M. Huang and B. Mutlu, "The repertoire of robot behavior: enabling robots to achieve interaction goals through social behavior," *J Hum-Robot Interact*, vol. 2, no. 2, pp. 80–102, Jun. 2013, doi: 10.5898/JHRI.2.2.Huang.

[31] L. McCallum and P. W. McOwan, "Face the Music and Glance: How Nonverbal Behaviour Aids Human Robot Relationships Based in Music," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '15. New York, NY, USA: Association for Computing Machinery, Mar. 2015, pp. 237–244. doi: 10.1145/2696454.2696477.

[32] C. T. Ishi, C. Liu, H. Ishiguro, and N. Hagita, "Head motions during dialogue speech and nod timing control in humanoid robots," in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, in HRI '10. Osaka, Japan: IEEE Press, Mar. 2010, pp. 293–300.

[33] C. Liu, C. T. Ishi, H. Ishiguro, and N. Hagita, "Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '12. New York, NY, USA: Association for Computing Machinery, 2012, pp. 285–292. doi: 10.1145/2157689.2157797.