"Wait, What Are You Guys Talking About?": Facilitating Bids for Conversation Entry using ConvoBuddy, A Chatbot Conversation Assistant

Sarah Pratt
Simon Fraser University

Kenny Zhang
Simon Fraser University

ACM Reference Format:
Sarah Pratt and Kenny Zhang. 2024. "Wait, What Are You Guys Talking About?": Facilitating Bids for Conversation Entry using ConvoBuddy, A Chatbot Conversation Assistant. In . ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Video Link:

1 ABSTRACT

Facilitating smooth entry into ongoing conversations is essential for fostering inclusive social interactions, yet it remains a challenging aspect of conversational dynamics. This paper presents ConvoBuddy, a novel conversational assistant designed to support users seeking to join active conversational groups based upon a preliminary needfinding user study and literature review. Using spatial orientation patterns (f-formations) to identify conversational clusters, ConvoBuddy leverages smartphone-based sensors to detect group structures in real time, mitigating challenges of occlusion and accessibility present in other sensing techniques. ConvoBuddy also incorporates a novel approach to multi-party conversational response generation, integrating entry timing and turn-taking management. The system aims to address user needs and reduce friction for those joining a conversation while minimizing disruptions for existing participants. This work contributes to the broader field of social computing by enhancing automated conversational technologies with socially adaptive, context-aware functions.

2 INTRODUCTION

Conversational groups represent one the most fundamental aspects of dynamic social interaction, making them a crucial component of relationship building, information seeking, identity formation, and social/professional networking [7].

Joining an ongoing conversation is a nuanced process that involves subtle social cues, spatial dynamics, and precise timing. Whether at a conference, a social gathering, or an informal meet-up, people frequently experience challenges when attempting to enter established conversational groups. For newcomers, this situation can provoke feelings of anxiety or apprehension, while existing group members may feel uncomfortable adjusting the flow or focus of their conversation. Conversational assistance tools can help alleviate anxiety, increase engagement, improve confidence, and enhance communication skills in the domain of social and language skills [22] [21]. However, current conversational technologies generally focus on one-on-one interactions or preset group conversations but lack functionality for dynamic group entry.

To address these issues, this study presents ConvoBuddy, a novel conversational assistant tool designed to facilitate smoother conversational group entry through a combination of sensor-driven group detection, turn-taking support, and conversational guidance. Central to the system is an f-formation detection mechanism, a concept from sociological research and social computing, which uses smartphone sensor data to infer spatial arrangements indicative of group boundaries. This approach allows for real-time identification of conversational groups without requiring external hardware, ensuring broad accessibility and adaptability.

Building on this foundation, the assistant employs a turn-taking detection and cueing system based on noise and silence patterns to identify opportune moments for conversational entry. These insights are coupled with integration with a large language model, ChatGPT, to dynamically summarize the ongoing conversation and provide evidence-based tips for initiating engagement. By aligning its suggestions with findings from social interaction literature, the system offers targeted guidance that balances contextual relevance with conversational norms.

The main contributions of this paper consist of the following:

- A user needfinding study involving 13 participants, provided insights into user needs and informing the design of the system.
- The ConvoBuddy system, a novel conversational assistant tool implemented as a web-based application that requires no additional hardware or software updates, making it easily deployable across diverse user bases.
- Discussion of findings and potential impact on conversational technology and human interaction dynamics for future implementations.

3 BACKGROUND AND RELATED WORKS

In settings with multiple free-standing groups, like networking events, conferences, or social gatherings, entering and exiting a conversational group can cause friction as group members may need to temporarily adjust the flow of conversation and their spatial arrangement to accommodate the entering or exiting party. Joining a conversation as a third-party is typically preceded by milling about, casual eavesdropping, and attempts to interrupt [10]. These are all moments which can provoke anxiety and distress in those

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

seeking to join a conversation as well as those who are actively engaged in focused conversation [20].

Our research aims to create a novel user interaction that facilitates smoother entry into ongoing multi-party conversations. To achieve this, we investigate three areas: the formation and detection of conversational groups, the mechanics involved in making a conversational bid through turn-taking, and existing systems aimed at improving conversation through technology.

3.1 Conversational Group Detection

In order to join a conversational group, first we must identify them. Groups engaged in focused interaction, such as a conversation, tend to arrange themselves in spatially oriented patterns, known as the *facing formation* or f-formation in sociology literature [14]. The f-formation framework is a widely used in the field of computer vision and social robots to facilitate socially appropriate physical orientation for a robot [4], this work inspires us to apply these findings to assist humans in approaching conversational groups and reduce the cognitive load and anxiety of the approaching party.

The f-formation consists of an arrangement in which people maintain a convex space (termed the o-space) to which everyone in the group has direct and equal access, although there can be different sub-types of the f-formation depending on the social dynamics and the number of people [14]. The purpose of the f-formation, though occurring through spontaneous and often subconscious cues and actions, is theorized to create a shared space exclusive of the world around the participants, in order to maintain a sense of shared reciprocation, engagement, and focus [14].

Automated detection of f-formations can serve as a proxy for identifying conversational groups. Prior work in computer vision has successfully detected these formations in still images [18]. Other approaches have leveraged automated f-formation detection for human-robot interaction, enabling social robots to adopt more socially acceptable navigation and orientation mannerisms through the use of cameras and motion sensors to detect these groups [1] [23] [5]. One limitation encountered in the approach, however, is occlusion and obstruction of the visual sensing device [3]. Our research aims to estimate f-formation detection by using sensor information from each users' smartphone. In this way, occlusion is no longer an issue, and it uses readily available devices that require no external equipment.

3.2 Turn-Taking

Turn-taking as a basic system of organizing a conversation was initially formulated by Sacks et. al. [17] and has since formed the basis of many conversational analysis techniques. Though ubiquitous throughout every conversation, turn-taking is a finely coordinated activity. The rhythmic quality of turn-taking is so predictable that it was used to train a neural network to extract a multi-speaker conversation from a mix of irrelevant speakers based on acoustic information alone [6].

To coordinate turns, turn-allocation techniques are used in which a current speaker may select a next speaker or parties may self-select [17]. Prior research has identified a wide range of verbal and non-verbal cues that signal turn-taking, including prosody, breathing patterns, eye gaze, gestures, and timing [19] [16]. While previous research has explored turn-taking cues, much of it assumes an ongoing conversation between known participants [8][9]. We seek to extend this work by focusing on turn-taking in the context of entry into a conversation, where the assistant must coordinate its suggestions for when and how to interject. Our implementation draws influence from the findings of Aoki et. al. in which they outlined four most successful (i.e., produced the most affiliative/affirmative responses) types of bids for new "floors", or new initiations of a conversation, within an ongoing multi-party conversation, in this case termed a "schism" [2]:

1. *Schisming by Schism-Inducing Turn (SIT)* - causes a change in topic, directly implicates a response often from a specific recipient or recipients. SITs often employ the usual resources of maximizing the chance of a successful bid for entry, such as timing the utterance to minimize speech overlap from interfering speakers the use of direct verbal address (e.g., names and pronouns), employment of gaze and physical gesture.

2. *Schisming by Toss-out* - like SIT, is topic-relevant to the in-progress conversation, occurring within its turn-taking system, and does not directly target a specific recipient or recipients, but is an open invitation for response.

3. *Schisming by Aside* - like toss-outs, are topic-relevant for the current conversation and do not strongly implicate a response. Asides can be positioned in overlap with the on-going conversation, and are not a time-sensitive as the previous two techniques. Asides are produced in a subdued voice

4. S*chisming by Retro-Sequence* - can occur when a turn that was not initially designed to initiate a new conversation is later treated as such by subsequent turns. Retro-sequences involve turns that acquire a different status than their initial one based on how others respond to them, i.e. laughing at something that wasn't intended to be funny.

We incorporate each of these techniques as prompts for suggestions for different types of bids for entry into an ongoing conversation.

3.3 Conversational Assistance Tools

Previous attempts at automated conversational assistance include the Smart Replies feature of Google's messaging application Allo, in which users expressed frustration toward the feature due to the limited variety of potential replies which did not capture a wide enough spectrum of genuine user preferences [11]. We incorporate this finding into our research by ensuring that conversational tips suggest multiple possible options and techniques, allowing the user to choose one that aligns with their preferred style of response.

Previous work has also explored automated social skills training which provides a wide variety of feedback on the user's social skills such as words per minute, voice pitch variation, pauses, filler word usage, and many more [22] [12], which is extremely beneficial for individuals practicing speaking skills on their own. Previous work has also gained purchase on conversational agents as facilitators of human-human interaction [13] [15]. However, one limitation of these prior works is the assumption that the user is already a member of an ongoing conversation. To negotiate entry into an ongoing conversation, we leverage a large-language model conversational assistant that is able to engage in interactions such that it can make helpful suggestions to a user seeking to join an ongoing conversation.

4 STUDY DESIGN

To inform the design of ConvoBuddy, we conducted a user need-finding survey to explore user anxieties related to conversational group entry as well as desires and expectations for a conversation entry support tool. Thirteen participants were recruited via social networks and online communities, with ages ranging between between 18 and 54, with the median age being between 18-24. The survey consisted of 9 free-response and multiple choice questions that took an average of about 5 minutes to complete. The survey included open-ended and multiple-choice questions designed to identify common challenges and expectations for a conversational entry support tool. Below, we present the key findings and highlight the trends that directly influenced the design of the system.

5 RESULTS

A majority of participants reported discomfort when joining ongoing group conversations. Over 61% (8 out of 13) described themselves as either "Uncomfortable" (46.15%) or "Very Uncomfortable" (15.38%). Only a small fraction felt "Comfortable" (15.38%), and none felt "Very Comfortable." This finding underscores the motivating necessity of a tool to support confidence and provide guidance during group entry.

When asked about any specific factors participants look for when deciding to join an ongoing conversational group, a few key themes emerged. Many participants emphasized the importance of aligning their knowledge or interests with the ongoing topic of discussion (e.g., "*I decide whether or not I have something to say based on what they're discussing*"). Participants also noted the importance of assessing conversational flow and identifying pauses before interjecting (e.g., "*When everyone stops talking, I will interrupt to ask my question*"). Some participants also expressed preference for joining groups where someone they know was already involved or when groups exhibited open body language (e.g., "*I'd feel most comfortable if someone I know was participating*").

These insights informed features such as conversational topic summarization, timing cues, and f-formation detection. We also acknowledge the importance of the physical presence of a familiar face when deciding to join a conversation, however, we do not focus on that as a feature of our conversational assistant because our conception of the tool is to be used at an event where there is limited pre-existing social context.

When asked about what factors would make group entry easier, participants expressed a strong preference for social and contextual cues to facilitate entry. Knowing ahead of time what the group is talking about before joining (84.62% of responses) was the most popular response, reinforcing the decision to include topic summarization.

When asked about what features would be most valued in a conversational support tool, the most frequently suggested answer were: notifying relevant conversations (61.54%) and suggesting content for contributions (61.54%), underscoring the need for intelligent summarization and recommendation systems, as well as timing cues to signal when to speak (61.54%) was also a high priority,

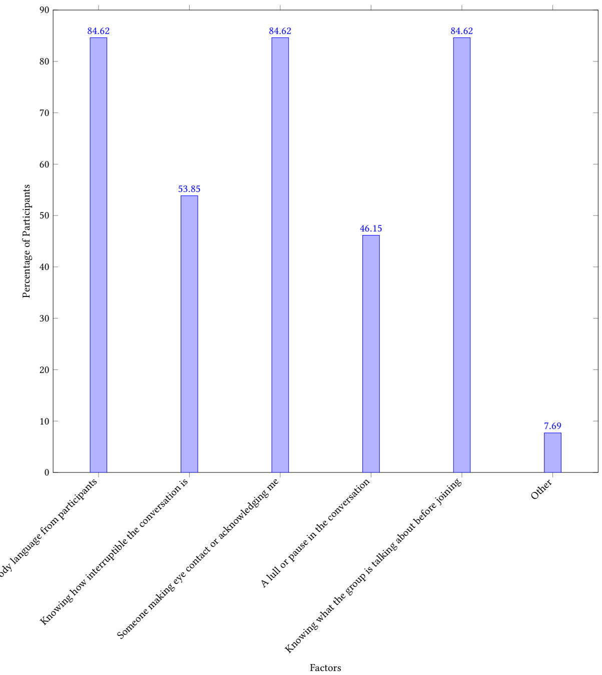

Figure 1: Participant responses to the question "What factors would make it easier for you to join a group conversation?"

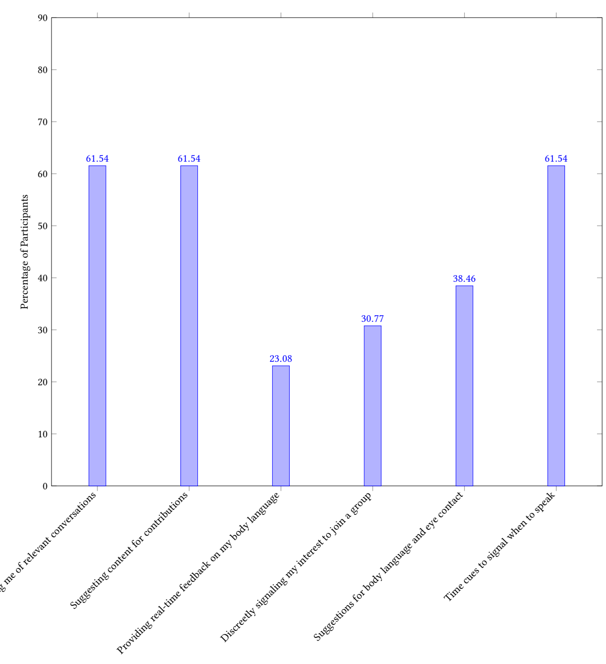

Figure 2: Participant responses to the question "If you had a technology-based tool to help you join a conversational group, how would you prefer this tool to assist you in joining conversations?"

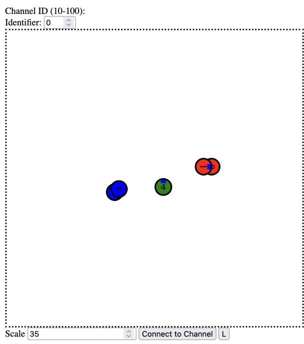

Figure 3: ConvoBuddy map visualization displaying the existing conversational groups in the networking space

confirming the need for real-time silence/noise detection. As such, we have incorporated all three of the most popular suggestions as a feature in the ensuing implementation.

6 IMPLEMENTATION

The conversational assistant was implemented as a mobile-friendly web application, consisting of a Python-based server for backend processing, a Node.js GPT-4o wrapper server and a Node.js-based application for real-time interaction and visualization, using the micro-service architecture framework. The decision to create a web application was motivated by the desire to create a tool that does not require external hardware or software installation, ensuring broad accessibility and eliminating technological barriers for users. First we will outline the intended user interaction process, then we will review the implementation details.

6.1 User Scenario

A user named William attends a networking event where there is limited pre-existing social context. In this scenario, we envision that the participants of the event all agree to use this conversational assistant tool to help with socializing. We assume that all participants have the web application ConvoBuddy open on their phones. We also assume that only people in the same physical space will access this web application.

When William navigates to ConvoBuddy (in this instance you can click ConvoBuddy), he can see a visualization of the existing conversational groups in the space, represented in Figure 3. Each user of ConvoBuddy is represented as a point on the map, with orientation denoted with an arrow. When conversational groups between users are detected through proximity and orientation, each point on the map is updated to have a distinguishing color. The colors help William correlate the physical layout of the conversational groups with the transcript summaries he is provided in the bottom half of the ConvoBuddy interface, see Figure 4. William scans the conversational summaries to select a conversation he is interested in joining.

Upon finding a topic he is interested in, in this case the red conversation, William opens the conversation to reveal a truncated transcript of the ongoing conversation, as well as some tips for joining, as displayed in Figure 5.

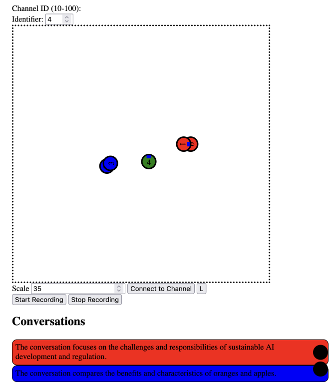

Figure 4: The conversational summaries are displayed in the bottom half of the user interface and are visually correlated via color to the conversational groups represented in the map.

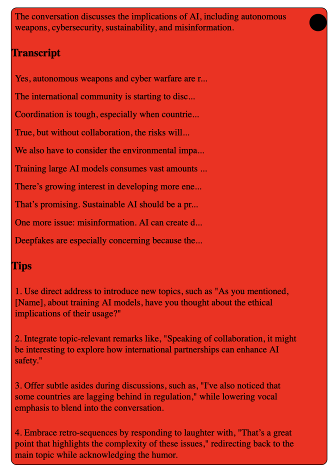

Figure 5: Conversational tips for joining an example conversation

Now that William knows the physical location and layout of the conversational group, the topic of conversation, and is equipped with some tips for joining, now he just needs to make his approach and take his turn. For this part of the user interaction, we have implemented a timing cue such that conversational silence occurs for a duration exceeding two seconds, the conversational tips box, as displayed in Figure 5, will turn white. This provides a subtle motivational nudge to William to take his turn to speak.

6.2 Backend Servers

There are two backend services that carry out the operations required: ConvoAnalyser, TranslateServer.

ConvoAnalyser is a Python server that is responsible for handling data processing, coordinating external API calls and relaying results to the frontend client. Key functionalities include F-formation detection using phone sensor data such as the built-in gyroscope and accelerometer passed in from the frontend, the server calculates a user's spatial movements and uses that to cluster and estimate conversational groups in real-time. Relative positioning mitigates the precision limitations of absolute GPS. Determining every user's absolute position in the room would require external hardware and sensors, which was not our aim. A basic clustering approach groups users based on relative proximity and orientation. Specifically, distances below a predefined threshold and aligned orientations within ±45° are considered indicative of an f-formation. This simple algorithm balances computational efficiency and sufficient accuracy for detecting f-formations in typical social settings.

It then processes speech transcript data to format conversation transcripts for summarization via an external LLM (ChatGPT 4o). The server then makes API calls to the LLM service to generate summaries of conversation on a per-group basis, and subsequently provide topic suggestions for the user seeking to join. The prompt was influenced by the prior work of A oki et. al. and formats the conversational entry tips such that they produce easy and quick to read tips for each of the four schisming strategies.

This is paired with TranslateServer, a Node.js server that acts as a wrapper for the GPT-4o API. It handles all the authentication and other non-essential bulk for requests and allows for ConvoAnalyser to easily make API requests with a simple POST request.

6.3 Node.js Application

The Node.js application interfaces with the Python server, receives sensor data, and provides real-time feedback to users. Its main functionalities include: using sensor data from phones to process relative spatial locations, with the assumption that all phones are initialized in the same reference frame. This approach addresses the limitations of GPS for precise absolute positioning.

The application recognizes speech using Web Speech API built-in speech-to-text service, and then maps speakers to their relative locations, and groups them into visually distinct conversational clusters in real time. Groups are color-coded for easy differentiation. A dynamic interface renders a spatial image of the conversation space, showing the user the location and structure of each group. A two-second silence/noise detector identifies pauses in the conversation, providing a visual cue to users. Despite latency limitations, this feature serves as an anticipatory signal to indicate optimal moments to join a conversation, rather than precise cues similar to that of an orchestra conductor. The application then displays real-time conversational tips, including topic suggestions, summaries of ongoing discussions, and timing cues to enhance user confidence and encourage participation. Leveraging an LLM to summarize conversations and suggest topics fulfills the user desire for context-aware assistance, directly addressing anxieties about conversational relevance.

7 DISCUSSION AND LIMITATIONS

The conversational assistant successfully demonstrates the integration of multiple modalities—spatial visualization, real-time conversational analytics, and AI-driven content generation—to support conversational group entry, as motivated by a user needfinding study. From an HCI perspective, the system highlights the potential of smartphone sensors and AI tools in mitigating social anxieties and encouraging more inclusive interactions. For developers and designers, our work showcases the possibilities of specialized conversational support tools without using any specialized hardware or having to install software, with the aim of maximizing widespread usability.

Phone Sensor Data Accuracy: Sensor data processed through a web app showed inconsistencies due to device variability and environmental factors, affecting the precision of spatial visualization. Latency Issues: API calls to the LLM occasionally introduced delays, reducing the responsiveness of timing cues. Initialization Requirement: The lack of absolute positioning necessitated manual initialization of phones in the same location, which could be cumbersome for users. Speech Transcription Quality: The speech recognition service struggled with accuracy in noisy environments, limiting the reliability of conversation summaries.

Robustness to Real-World Conditions: Systems relying on smartphone sensors and speech transcription must account for environmental variability, such as noise and user movement.

8 LIMITATIONS AND FUTURE WORK

There is room for improvement when it comes to improved sensor integration. Implementations with newer phones could leverage higher-resolution location tracking technologies such as Ultra-Wideband (UWB). This would likely improve the reliance and accuracy of speaker location tracking. We could also employ machine learning techniques to improve speaker diarization to remove the need for users to each carry their own microphone (in this case their phone microphones). There is also room for improvement when it come to latency and the improvement of finer-tuned timing cues. We were limited by the latency between server calls and thus our timing cues were rather coarse-grained. Using on-device processing could potentially improve the performance of the timing cues feature. A more rigorous technical study could be done that compares the detected conversational groups with real-life conversational groups to determine accuracy metrics.

Future efforts could also include structured evaluations to conduct more rigorous user studies, especially diverse cultural and social contexts as well as longer time horizon studies. It would be helpful to study this tool within an array of diverse social contexts for which conversation entry tips could be adapted. It would also be useful to conduct a formal user evaluation of the prototype from a user experience perspective. For instance, we could gather our participants from the needfinding survey and allow them to use this tool in a mixed group with each other to see how intuitive and useful the tool is. We could iterate upon this by incorporating feedback into the design at each stage of the design process.

Future work could also adopt a more user-centric approach that incorporates user preferences ahead of time, for instance, the preference to join smaller groups or a notification about conversations taking place that concern a certain preferred topic.

9 CONCLUSION

The development of ConvoBuddy, a conversational assistant designed to facilitate smoother entry into conversational groups, provides key contributions to the field of human-computer interaction and human-AI interaction. By integrating conversational group, or f-formation detection, turn-taking cues, and AI-driven content generation, the tool demonstrates a novel approach to mitigating social anxieties and fostering inclusivity in group interactions. Its implementation as a web-based application requiring no specialized hardware ensures convenience and scalability across diverse user contexts.

This work is significant for its focus on dynamic, real-time group conversations and entry strategies—a domain often overlooked by existing conversational technologies, which typically center on one-on-one or predefined group interactions. By addressing user needs identified through a comprehensive needfinding study, the system highlights how AI and sensor-driven interfaces can empower individuals to navigate complex social environments confidently.

The insights gained from this research underscore the importance of designing conversational tools that are adaptive, context-aware, and inclusive. ConvoBuddy's potential applications, from networking events to informal gatherings, illustrate the transformative impact of technology in enhancing human interactions. Future work will build on these findings, refining technical robustness and exploring broader cultural and linguistic contexts to maximize the tool's applicability.

ACKNOWLEDGMENTS

We thank all the participants who dedicated their time and effort to this study.

REFERENCES

[1] P. Althaus, H. Ishiguro, T. Kanda, T. Miyashita, and H.I. Christensen. 2004. Navigation for human-robot interaction tasks. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, Vol. 2. 1894–1900 Vol.2. https://doi.org/10.1109/ROBOT.2004.1308100 ISSN: 1050-4729.

[2] Paul M. Aoki, Margaret H. Szymanski, Luke Plurkowski, James D. Thornton, Allison Woodruff, and Weilie Yi. 2006. Where's the "party" in "multi-party"? analyzing the structure of small-group sociable talk. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work (CSCW '06)*. Association for Computing Machinery, New York, NY, USA, 393–402. https://doi.org/10.1145/1180875.1180934

[3] Hrishav Bakul Barua, Theint Haythi Mg, Pradip Pramanick, and Chayan Sarkar. 2021. Detecting socially interacting groups using f-formation: A survey of taxonomy, methods, datasets, applications, challenges, and future research directions. https://arxiv.org/abs/2108.06181v2

[4] Hrishav Bakul Barua, Theint Haythi Mg, Pradip Pramanick, and Chayan Sarkar. 2024. Enabling Social Robots to Perceive and Join Socially Interacting Groups Using F-formation: A Comprehensive Overview. *J. Hum.-Robot Interact.* 13, 4 (Oct. 2024), 48:1–48:48. https://doi.org/10.1145/3682072

[5] Hrishav Bakul Barua, Pradip Pramanick, Chayan Sarkar, and Theint Haythi Mg. 2020. Let me join you! Real-time F-formation recognition by a socially aware robot. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 371–377. https://doi.org/10.1109/RO-MAN47096.2020.9223469 ISSN: 1944-9437.

[6] Tuochao Chen, Qirui Wang, Bohan Wu, Malek Itani, Emre Sefik Eskimez, Takuya Yoshioka, and Shyamnath Gollakota. 2024. Target conversation extraction: Source separation using turn-taking dynamics. , 3550–3554 pages. https://doi.org/10.21437/interspeech.2024-225

[7] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300705

[8] Erik Ekstedt and Gabriel Skantze. 2020. TurnGPT: a Transformer-based Language Model for Predicting Turn-taking in Spoken Dialog. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 2981–2990. https://doi.org/10.18653/v1/2020.findings-emnlp.268

[9] David R. Gibson. 2010. Marking the Turn: Obligation, Engagement, and Alienation in Group Discussions. *Social Psychology Quarterly* 73, 2 (June 2010), 132–151. https://doi.org/10.1177/0190272510371456 Publisher: SAGE Publications Inc.

[10] Erving Goffman. 1979. *Footing*. Vol. 25. De Gruyter. 1–30 pages. https://doi.org/10.1515/semi.1979.25.1-2.1 Publisher: De Gruyter Mouton Section: Semiotica.

[11] Jess Hohenstein and Malte Jung. 2018. AI-Supported Messaging: An Investigation of Human-Human Text Conversation with AI Support. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188487

[12] Mohammed Ehsan Hoque. 2013. *Computers to help with conversations : affective framework to enhance human nonverbal skills*. Thesis. Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/91845 Accepted: 2014-11-24T18:39:24Z.

[13] Katherine Isbister, Hideyuki Nakanishi, Toru Ishida, and Cliff Nass. 2000. Helper agent: designing an assistant for human-human interaction in a virtual meeting space. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '00)*. Association for Computing Machinery, New York, NY, USA, 57–64. https://doi.org/10.1145/332040.332407

[14] Adam Kendon. 2009. Spacing and orientation in co-present interaction. In *Proceedings of the Second international conference on Development of Multimodal Interfaces: active Listening and Synchrony (COST'09)*. Springer-Verlag, Berlin, Heidelberg, 1–15. https://doi.org/10.1007/978-3-642-12397-9_1

[15] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joonhwan Lee. 2020. Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376785

[16] Stephen C. Levinson and Francisco Torreira. 2015. Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology* 6 (2015), 731. https://doi.org/10.3389/fpsyg.2015.00731

[17] HARVEY Sacks, EMANUEL A. Schegloff, and GAIL Jefferson. 1978. A Simplest Systematics for the Organization of Turn Taking for Conversation*. In *Studies in the Organization of Conversational Interaction*, JIM Schenkein (Ed.). Academic Press, 7–55. https://doi.org/10.1016/B978-0-12-623550-0.50008-2

[18] Francesco Setti, Chris Russell, Chiara Bassetti, and Marco Cristani. 2015. F-Formation Detection: Individuating Free-Standing Conversational Groups in Images. *PLOS ONE* 10, 5 (May 2015), e0123783. https://doi.org/10.1371/journal.pone.0123783 Publisher: Public Library of Science.

[19] Gabriel Skantze. 2021. Turn-taking in Conversational Systems and Human-Robot Interaction: A Review. *Computer Speech & Language* 67 (May 2021), 101178. https://doi.org/10.1016/j.csl.2020.101178

[20] Avelie Stuart, Clifford Stevenson, Miriam Koschate, Jessica Cohen, and Mark Levine. 2022. 'Oh no, not a group!' The factors that lonely or isolated people report as barriers to joining groups for health and well-being. *British Journal of Health Psychology* 27, 1 (2022), 179–193. https://doi.org/10.1111/bjhp.12536 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bjhp.12536.

[21] Tzu-Yu Tai. 2024. Comparing the effects of intelligent personal assistant-human and human-human interactions on EFL learners' willingness to communicate beyond the classroom. *Comput. Educ.* 210, C (March 2024). https://doi.org/10.1016/j.compedu.2023.104965

[22] Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Tomoki Toda, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura. 2015. Automated Social Skills Trainer. In *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*. Association for Computing Machinery, New York, NY, USA, 17–27. https://doi.org/10.1145/2678025.2701368

[23] Shih-An Yang, Edwinn Gamborino, Chun-Tang Yang, and Li-Chen Fu. 2017. A study on the social acceptance of a robot in a multi-human interaction using an F-formation based motion model. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2766–2771. https://doi.org/10.1109/IROS.2017.8206105 ISSN: 2153-0866.