Robots Who Interrupt Talk in Meetings

Olli Niinivaara
olli.niinivaara@helsinki.fi
University of Helsinki
Department of Practical Philosophy
Helsinki, Finland

ABSTRACT

Knowledge sharing is an important aspect in most meetings. Personal characteristics of some participants, such as their (in)ability or (un)willingness to take the floor, may have a negative effect on the quality of knowledge sharing; some people tend to talk too much, while others have difficulties in making themselves heard. A robotic facilitator can be used to distribute the floor time more efficiently. While current research is mostly focused on encouraging participants to talk, this paper suggests interruption functionality to discourage speakers from talking. The facilitator gathers turn-taking signals from the participants and expresses them on their behalf. It hides the identity of individuals, making it easier for everyone to take action. The facilitator represents the signals coherently for all signalers, which equalizes the differences in social signalling skills, and makes it easier for the speaker to interpret the signals. It continuously gathers feedback from all participants, and thereby can represent the collective mood of the audience and smooth out outlier reactions. The facilitator can be programmed to act in a germane, courteous and attentive manner, which helps keeping the meeting mood high.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative interaction**; Collaborative and social computing systems and tools; • **Computer systems organization** → *External interfaces for robotics*; • **Social and professional topics** → User characteristics.

KEYWORDS

conversational agent, automated facilitation, backchannel, interruption, turn-taking, collective feedback, nonverbal communication

> If you had to identify, in one word, the reason why the human race has not achieved, and never will achieve, its full potential, that word would be "meetings".
>
> Dave Barry

1 INTRODUCTION

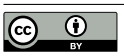
Meetings are central to organizational life, and the time people spend in meetings is ever-increasing [1, 12]. To ensure the efficiency of a meeting, one important aspect is to give everyone an opportunity to speak out. The more participants there are, and the less familiar the participants are with each other, the more difficult the orchestration of turn-taking becomes. A skilled meeting facilitator (moderator, mediator, chair, host) definitely helps, but sometimes such a person is not available. And even then, the amount of participants can make facilitating just too hard for humans. Therefore, it makes sense to research various technological tools that could help in running meetings more efficiently. In words of Woolley et al. [23]: "Could a group's collective intelligence be increased by, for example, better electronic collaboration tools?"

Maybe the most sophisticated idea for orchestrating group conversations is the use of automated facilitators [see e.g. 2, 3, 11, 16, 17]. However, existing systems are principally designed to distribute floor time evenly by encouraging the more silent listeners to speak. Such an approach has two shortcomings. One is that in order to speak, there has to be an opportunity to do that. It is not uncommon that a small loud and dominant minority (or even one person, say "a besserwisser on a mission") can steal the show and render a meeting quite inefficient. Therefore encouragement to speak may not be enough in itself, but a speaker should sometimes be encouraged *not* to speak, or even be interrupted. The other shortcoming is that distributing floor time evenly is not a goal in itself. It is namely fine that some participants talk more than others, if they are the ones who have more important information to communicate. It would be counterproductive to force those participants to speak who do not happen to have anything interesting to say.

As an addition to automated meeting facilitators, I suggest interruption functionality that is designed specifically for discouraging the current speaker from talking. A recent review on turn-taking in conversational systems [18] indicates that such a functionality has not yet been widely considered. As exceptions, Bergstrom & Karahalios [3] and Bohus & Horvitz [4] have considered the possibility of such negative feedback. However, both ended up not implementing such a functionality, and therefore do not discuss it further.

The most relevant related work is a robot for managing turn-taking by Gonnot at al. [7], which is also a system designed to anonymously collect and then non-verbally express various negative feedback signals. However, their input interface design requires somewhat cognitive processing, which might be distracting. As an alternative to that, I suggest a more minimal input interface. More problematically, their output interface is based on robotic tractor Cozmo[1], that communicates by steering around a meeting table. This communication media is quite unnatural and obviously does not scale to larger conferences. Instead, I suggest that a humanoid robot is used to express the social cues, which should be much easier to interpret. And, above all, Gonnot at al. avoid interrupting a talk, whereas here it is one of the main objectives.

[1] https://www.digitaldreamlabs.com/pages/meet-cozmo



HRI 2023 Workshop on Human-Robot Conversational Interaction, March 13, 2023, Stockholm, SE

2 INPUT

The most important UX input requirements in this context seem to be accuracy, user equality and discretion. Accuracy is required, because discouraging speaking is a delicate matter, and interrupting someone is a drastic act. An erroneous operation here would easily render a conversational agent detrimental. User equality means that floor time should not depend on the personal characteristics of the participants, such as on their degree of glossophobia, extraversion, their social skills in expressing backchanneling signals, or their abilities to do surface acting. Discretion is required so that even the most shy and socially anxious listeners feel safe using the system. An anonymous and indiscernible input interface also avoids distracting the ongoing discussion, so that the decision on whether and how to act can be controlled by the system.

Systems that can detect the emotional intent from nonverbal backchanneling cues are currently being developed (e.g. [2, 4, 5, 11]). However, such an input interface would not fulfill the requirements for user equality and discretion, because they require the users to be brave enough to stand out, and to stand out in a way that is correctly interpreted by the detection system. Therefore I suggest it is better to let the participants use their own device (laptop or smartphone[2]) to signal their intents. The participants could, for example, surf to a web address given in the start of meeting, and use the system by clicking on a set of buttons (see below) offered there.

To decide the most useful set of commands, experimental work is needed. While a lot of research on turn-taking and conversational agents exists, research that explicitly concerns interruptions and other acts that discourage speech is scarce. Seminal work on the area is Lycan [10]. Goldberg [6] gives a typology of interruptions but focuses on vocal utterances. Schlöder [14] presents a taxonomy of rejection moves in dialogue. Gonnot et al. [7] is an implementation with a command palette with many useful negative feedback functionalities.

Based on the aforementioned previous work, the following action set is suggested as a starting point. The suggested actions are divided in three categories. Category **Advice** contains impressions signaling that the speaker may continue, but should somehow adjust the presentation. Category **Comment** contains signals stating that the speaker should give the floor for a short moment, but may then continue. Moreover, these comment signals come in two moods: stating either that an intermediate commentary in general would be in place, or that the signaler itself wants to utter this kind of a comment. Finally, category **Stop** contains ways to signal the speaker to give the floor for now.

Advices:

🤔 **Explain**: ”We did not understand. Please explain with more detail”

🤨 **Doubtful**: ”We found that hard to believe. Please be more convincing”

😴 **Skip**: ”You are wasting our time. Please state your point on this topic”

Comments:

🕵️ **Questionable**: ”Let me/us ask you a question”

🤦 **Mistake**: ”Let me/us correct you”

💁 **Dialogue**: ”Let me/us answer that”

🗣️ **Announcement**: ”I have/there is a short announcement to make”

Stops:

🤦 **Inappropriate**: ”Your delivery does not belong here”

⌛ **Overtime**: ”Your time is up”

🤼 **Dispute**: ”You are just arguing with each other, please respect our time and continue that somewhere else”

🤐 **Secret**: ”You cannot talk about that here”

In a fully working system, some auxiliary actions would probably also be required. Some actions could be directed towards the facilitator. For example, a **Cancel** signal could be used to cease the current robot script when it has become irrelevant due to later happenings. Some actions could be directed toward other audience members. For example, a **Calm down** signal could be used to pacify a restless part of audience.

As an extension, the signals could come in degrees. A weak signal would mean that the signaler is in doubt, so that the output script should be started only if some majority of listerners also entertains the same opinion (”I wonder if there was a mistake”). A strong signal would mean that the signaler wants the signal to be carried out in full force, irrespective of what other opinions are currently active in the audience (”I have an announcement to make: the building is on fire!”).

3 OUTPUT

The most important UX output requirements in this context seem to be non-disruptiveness and effectiveness. Non-disruptiveness means that signals should not confuse or agitate the speaker. Effectiveness means that it should not be possible for the speaker to ignore the signals, either deliberatively or by accident.

The aforementioned requirements contradict each other. This dilemma can be solved by using gradual, nonverbal output. Graduality means that an output action script starts with quite inconspicuous signals like small gestures, and progresses towards more intrusive measures, like speaking aloud. Nonverbal output can be implement by using a humanoid (antropomorphic) output UI. Human gestures will be effortlessly and immediately recognized and emotional response to them is involuntary, without any higher cognitive processing required [22]. Urakami & Seaborn [21] and Saunderson & Nejat [13] give reviews on how robots can influence humans with such nonverbal communication.

The operation and the granularity of output is explained next with scenarios[3].

[2] there a also Audience Response Systems [8] for this purpose, see https://www.rn.inf.tu-dresden.de/arselector/ for list

[3] Note on format: Emotional responses, such as *gets bored* are denoted in italics; Audience input signals (”button clicks”), such as **Mistake** are denoted in bold

scenario 1

Let us imagine a big scientific conference where a speaker named Amy has just made a serious mistake that should be immediately corrected.

LISTENER A: (*recognizes a potential mistake*): **Mistake**
ROBOT: blinks eyes, slightly jerks head
MORE LISTENERS: (*also start to suspect a problem*): **Mistake**
ROBOT: squeaks eyebrows, rubs chin
MORE LISTENERS: (*agree that something is not right*): **Mistake**
ROBOT: scratches ear, shakes head, says: "hmmmm..."
AMY: (*decides to react*): Says: "If you spotted something, please help me out"
POTENTIAL COMMENTATORS: (*willing to step up*): **Let me answer that**
ROBOT: raises hand, stares at the speaker
AMY: (*recognizes that someone is willing to comment*): Interrupts the speech
ROBOT: (*recognizes that speaker is silent*): Says: "We have a comment from the audience". Lowers hand. (*Selects one potential commentator to get the floor*)
COMMENTATOR: (*Notices that robot has chosen her*): Explains the mistake
OTHER POTENTIAL COMMENTATORS: (*Find the explanation adequate*): Remove their **Let me answer that** -signals
AMY: (*Notes that incident is solved*) Continues the presentation

scenario 2

Let us imagine a town community public meeting (a meeting where public can discuss with officials), where one citizen (Bob) is ranting about an irrelevant topic.

BOB: keeps on talking
LISTENER A: (*gets bored*) **Skip**
ROBOT: sighs, drums fingers
BOB: continues
SOME MORE LISTENERS: (*get bored*) **Skip**
ROBOT: yawns and stretches, gesticulates as if looking at its wrist watch
LISTENER B: (*gets annoyed*) **Inappropriate**
ROBOT: scratches forehead, shakes head, rolls eyes
BOB: starts to speak even more aggressively
SOME LISTENERS: (*Hope someone to answer to Karen*) **Let's Dialogue**
ROBOT: raises hand, sweeps hand toward audience
SOMEONE: (*Ready to stop Bob*) **I have an Announcement**
ROBOT: stands up, points toward audience, coughs loudly
BOB: Keeps on talking
MAJORITY OF LISTENERS: (*Hoping Bob to shut up*) **Inappropriate** or **Skip** or **There is an announcement**
ROBOT: walks toward Bob, raises both hands, says: "Please, stop speaking now, or I start singing loudly."
BOB: shuts up.
ROBOT: signals to the announce-maker to go on, walks back to its chair and sits down

4 ETHICAL CONSIDERATIONS

Let us, for argument's sake, assume robotic facilitators to be a product with huge sales potential. A technology that deeply disrupts how people communicate with each other may bring benefits, but might also cause some unintended negative consequences. Therefore it is justifiable to consider also the potential negative effects before such robots are being deployed in massive scale.

Maybe the most pressing issue in our times is achieving environmental sustainability, and mass-scale production of robots would be detrimental due to their carbon footprint and use of rare-earth elements in actuators, batteries and motherboards. Up to a point, animated virtual avatar facilitators could be used instead of robots. However, they might be far less effective interruptors than physical robots. To reduce the amount of robots, a common application programming interface for robots from different manufacturers would enable one robot to be used in multiple roles, meeting facilitation included. Robots could also be shared or rented between users, for example borrowed from libraries. A single robot could serve multiple meetings by walking from meeting to meeting. But if we assume that a robot is already manufactured, adding the interruption functionality to it would not cause further environmental damage in itself, especially if the interruption component decreases meeting times and thereby energy consumption.

Participants who do not have suitable device at hand would be outcasts in robot-facilitated meetings. While it is realistic to assume that everyone can bring a laptop or a smartphone, oversights and malfunctions do happen. There are four remedies. One is to keep spare devices available by the meeting organizer. Another is to try to automatically detect the signals from the non-verbal (and verbal) cues of the participants. A third is to pair people without a device with those who have. A fourth is to not use a robot whenever someone in the audience so wishes. As different options to handle the situation exist, missing a device does not seem to be a critical obstacle.

An automated facilitator can record sensible data from meetings. Cloud video conferencing software are routinely used in meetings nowadays, hinting that electronic meeting tools in general do not pose a security risk to organizations. However, a system that collects quantitative data about meeting behavior poses an intra-organizational privacy risk. Especially, using this kind of data as performance indicators for workforce assessment might be tempting. Such data could reveal, for example, which people agree and disagree with each other the most. Using a facilitator robot as an intra-organizational surveillance spy should be discouraged or technically prevented (for example by deleting all data after every meeting). Otherwise even the awareness that every action is being recorded might incentivize participants to focus their energy on gaming the system instead of on making the meeting productive.

Continuous use of automated facilitators everywhere might lead to a situation in which people become quite bad at communicating, cooperating or coordinating on anything without a robot being present. It is self-evident that advances in technology change the skills people need. However, it is reasonable to ask, whether some basic communication skills are so essential to us that we should not let them degenerate. On the positive side, people are naturally quite good at imitation, and therefore watching a robot might teach

people how to facilitate well. People should also be encouraged to hold meetings without a robot every now and then, in order to keep old-fashioned meeting skills at some base level. However, our understanding of what happens when robots are placed within groups or teams for longer times is highly limited [15], therefore more research on this area is needed.

The user interface operations suggested here are based on how people currently orchestrate meetings. But of course also totally novel input and output operations could be created that have no basis in the current social reality. Such operations might turn out to be much more efficient than current ways of managing turn-taking. After all, current signals were never engineered, but are more or less ad hoc results of historical (evolutionary [22], cultural [9], and societal [e.g. 20]) developments. In our ancient past the individuals who were the strongest and most aggressive might have dominated meeting outcomes. Today, the most charismatic and socially skilled people have an advantage over the socially clumsy introverts. In the future, those who are the best using the technhologies to their advantage might dominate both the aggressive and the charismatic ones. Such a future would be ethically more acceptable than the status quo in the sense that it is arguably easier to learn to use meeting tools than to change one's personality into an extrovert. But the basic requirement is that everyone must have an equal opportunity to learn these meeting technologies.

5 CONCLUSIONS AND FUTURE WORK

Automated facilitators are an advanced technology for managing turn-taking in meetings. Current research has mostly focused on sharing the floor-time evenly by encouraging the passive participants to engage more. As an additional counterpart, a functionality to discourage the speaker from talking was suggested. The facilitator discreetly gathers reactions from the audience, and performs the collective feedback with gradual, polite cues that become harder and harder for the speaker to ignore. In this way, the opportunity to interrupt the speaker does not anymore depend on the personal characteristics or the social skills of an individual participant, making the meeting experience more equal to everyone. The speaker does not need to be aware and understand all the simultaneous signals from the audience, but can concentrate on the feedback by the facilitator.

The next research step is to implement an interruption-capable conversational agent and run experiments with it. However, implementing only the interruptive capabilities would probably not work, because such a robot could bring about a quite negative meeting atmosphere. To counterbalance, some rapport-building backchanneling functionality should also be implemented.

While most HRI experiments are lab experiments, field experiments seem more adequate here. The atmosphere and dynamics of a (heated) meeting may be hard to create synthetically. Besides, the novelty effects [19] of introducing a robot are probably much higher in group settings than in dyadic interaction. To let the novelty effects wear off, a group should use an automated facilitator in their meetings until it becomes a routine.

Anyway, designing experiments for measuring effects on large group behaviour seems challenging. This suggests that the scarcity of experimental research on conversational robots as members of large groups does not necessarily stem from uselessness of such research, but from the hardness of doing it.

ACKNOWLEDGMENTS

This work has been funded by NordForsk.

REFERENCES

[1] Joseph A. Allen and Nale Lehmann-Willenbrock. 2022. The key features of workplace meetings: Conceptualizing the why, how, and what of meetings at work. *Organizational Psychology Review* (Sept. 2022). https://doi.org/10.1177/20413866221129231

[2] David Ayllon, Ting-Shuo Chou, Adam King, and Yang Shen. 2021. Identification and Engagement of Passive Subjects in Multiparty Conversations by a Humanoid Robot. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 535–539. https://doi.org/10.1145/3434074.3447229

[3] Tony Bergstrom and Karrie Karahalios. 2009. Vote and Be Heard: Adding Back-Channel Signals to Social Mirrors. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I (INTERACT '09)*. Springer-Verlag, Berlin, Heidelberg, 546–559. https://doi.org/10.1007/978-3-642-03655-2_61

[4] Dan Bohus and Eric Horvitz. 2010. Facilitating multiparty dialog with gaze, gesture, and speech. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI '10)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/1891903.1891910

[5] Konstantinos Bousmalis, Marc Mehu, and Maja Pantic. 2013. Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools. *Image and Vision Computing* 31, 2 (Feb. 2013), 203–221. https://doi.org/10.1016/j.imavis.2012.07.003

[6] Julia A. Goldberg. 1990. Interrupting the discourse on interruptions: An analysis in terms of relationally neutral, power- and rapport-oriented acts. *Journal of Pragmatics* 14, 6 (Dec. 1990), 883–903. https://doi.org/10.1016/0378-2166(90)90045-F

[7] Alix Gonnot, Christine Michel, Jean-Charles Marty, and Amélie Cordier. 2019. Social Robot as an Awareness Tool to Help Regulate Collaboration. In *Workshop on Robots for Learning (R4L) at the 28th IEEE International Conference on Robot & Human Interactive Communication*. New Delhi, India. https://hal.archives-ouvertes.fr/hal-02317785

[8] Robin H. Kay and Ann LeSage. 2009. Examining the benefits and challenges of using audience response systems: A review of the literature. *Computers & Education* 53, 3 (Nov. 2009), 819–827. https://doi.org/10.1016/j.compedu.2009.05.001

[9] Han Z. Li. 2001. Cooperative and Intrusive Interruptions in Inter- and Intracultural Dyadic Discourse. *Journal of Language and Social Psychology* 20, 3 (Sept. 2001), 259–284. https://doi.org/10.1177/0261927X01020003001

[10] William G. Lycan. 1977. Conversation, politeness, and interruption. *Linguistics* 10, 1-2 (March 1977), 23–53. https://doi.org/10.1080/08351819709370438

[11] Yoichi Matsuyama, Iwao Akiba, Shinya Fujie, and Tetsunori Kobayashi. 2015. Four-participant group conversation: A facilitation robot controlling engagement density as the fourth participant. *Computer Speech & Language* 33, 1 (Sept. 2015), 1–24. https://doi.org/10.1016/j.csl.2014.12.001

[12] Joseph E. Mroz, Joseph A. Allen, Dana C. Verhoeven, and Marissa L. Shuffler. 2018. Do We Really Need Another Meeting? The Science of Workplace Meetings. *Current Directions in Psychological Science* 27, 6 (Dec. 2018), 484–491. https://doi.org/10.1177/0963721418776307

[13] Shane Saunderson and Goldie Nejat. 2019. How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human–Robot Interaction. *International Journal of Social Robotics* 11, 4 (Aug. 2019), 575–608. https://doi.org/10.1007/s12369-019-00523-0

[14] Julian J. Schlöder and Raquel Fernández. 2019. How to Reject What in Dialogue. In *Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue*. SemDial, London, United Kingdom. http://semdial.org/anthology/Z19-Schlder_semdial_0014.pdf

[15] Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F. Jung. 2020. Robots in Groups and Teams: A Literature Review. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 176:1–176:36. https://doi.org/10.1145/3415247

[16] Ameneh Shamekhi. 2020. *Conversational Agents For Automated Group Meeting Facilitation.* Ph.D. Thesis. Northeastern University. https://repository.library.northeastern.edu/files/neu:m046qf41d

[17] Gabriel Skantze. 2017. Predicting and Regulating Participation Equality in Human-robot Conversations: Effects of Age and Gender. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. Association for Computing Machinery, New York, NY, USA, 196–204. https://doi.org/10.1145/2909824.3020210

[18] 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

[19] Catharina Vesterager Smedegaard. 2019. Reframing the Role of Novelty within Social HRI: from Noise to Information. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 411–420. https://doi.org/10.1109/HRI.2019.8673219

[20] Lynn Smith-Lovin and Charles Brody. 1989. Interruptions in Group Discussions: The Effects of Gender and Group Composition. *American Sociological Review* 54, 3 (1989), 424–435. https://doi.org/10.2307/2095614

[21] Jacqueline Urakami and Katie Seaborn. 2022. Nonverbal Cues in Human-robot Interaction: A Communication Studies Perspective. *ACM Transactions on Human-Robot Interaction* (Dec. 2022). https://doi.org/10.1145/3570169

[22] Gerben A. van Kleef and Stéphane Côté. 2022. The Social Effects of Emotions. *Annual Review of Psychology* 73, 1 (2022), 629–658. https://doi.org/10.1146/annurev-psych-020821-010855

[23] Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science* 330, 6004 (Oct. 2010), 686–688. https://doi.org/10.1126/science.1193147