MediatorBot: Autonomous Mediator for Multiparty Discussions

Julia Booth, Ryan Dowse, Patrick Engelbert, Becks Simpson

***Abstract*— Meetings are conducted in companies as a way of encouraging the sharing of ideas among employees and agreeing business actions. In order to improve the efficiency of meetings mediators can be used to balance the participants discussion. However, human mediators are often not objective, being biased to a certain idea or decision. In this report we introduce the idea that a humanoid robot can be used as an unbiased and more efficient mediator.**

I. INTRODUCTION

In businesses around the world a large proportion of employees time is taken up with meetings. These meetings are integral to the running of the businesses, encouraging collaboration which is key to their success. Unfortunately, many of these meetings are unproductive and inefficient uses of employees time due to some members dominating the conversation and other members not providing their input.

Often the way to overcome this lack of balance in the discussion is to appoint a mediator, whose responsibility is to control the flow of the meeting and ensure that all participants are given an opportunity to speak. However, the mediator often evaluates their performance in balancing the discussion more highly than the other meeting participants [1], who can feel that the mediator was biased.

In this project we propose that a humanoid robot, MediatorBot, can act as a truly unbiased mediator and therefore will not only better accepted by the members of a meeting but will also be more efficient. MediatorBot will be able to balance the discussion equally among the meeting participants and alleviate the need for an extra employee to attend the meeting to act as a mediator, making the meeting more efficient for the business.

II. RESEARCH HYPOTHESIS

We propose the following hypotheses:

1) The robot mediator will conduct a more efficient meeting than a human mediator, ensuring that members participate equally to reach a conclusion.
2) The members of a meeting will be more satisfied with the robot as a mediator, feeling that it is unbiased.

III. BACKGROUND

In group discussion, members take on a multitude of functional roles[2]. Some of these roles cause some participants to take on a large share of the discussion while others contribute to a much lesser extent. This has the unfortunate side effect of reducing the meeting effectiveness[1] due to discussion being limited to only a few participants at a time with others not engaging with the conversation. To counter this, a mediator or facilitator can be used who will guide and focus the group.

A good mediator should be: "An individual who enables groups and organizations to work more effectively; to collaborate and achieve synergy. He or she is a 'content neutral' party who by not taking sides or expressing or advocating a point of view during the meeting, can advocate for fair, open, and inclusive procedures to accomplish the group's work" [3]. Based on this description, the mediator is a good match for an impartial robot that can judge the situation of the group and give speaking turns to those that are regularly drowned out or to stop the discussion if it goes off track.

Matsuyama et al.[4] proposed the use of a robot as a fourth participant for a three party discussion. This robot will listen to the conversation and act as an active participant, interjecting its own view into the conversation in an attempt to engage currently neglected participants with the robot and then each other. While this approach is useful in a smaller and informal setting, the role of a mediator is to guide the discussion and allow the group to reach a consensus without having their own prejudice or influencing the decision[5].

A system as proposed by Kim et al.[6] provides a passive system of monitoring the discussion and showing feedback on the status of the discussion on the users' mobile phones. This system showed success in changing the participants speaking habits during use, reducing the overlap time of multiple speakers down from 50.1% to 31.8%.

In a similar vein, Zhang et al.[7] expanded on this work to create a more complicated GUI to express both individual but also group dynamics back the the participants of the discussion. This uses the CALO Meeting Assistant[8] as a basis for its meeting understanding and allows it to track who spoke when, to whom and the topics that they talk about.

Both of these methods have the issue that they are abstract and provide feedback only when the participants request it. To expand on this, MediatorBot will provide a physical presence, similar to the proposal by Matsuyama et al.[4] which will be the contact point for the software with the users which will be able to provide commands to the participants and act as a passive mediator until its presence is required.

In addition, Microsoft is currently working on a similar research project [9] , providing all of the same features that our system is expected to do with several additions such as: Addressee Detection, Agreement/Disagreement detection and Decision Extraction. While all of these features would be nice to have in MediatorBot, this framework is not yet available so we must design our system to include only the necessary of these features for our purpose.

IV. DESIGN

A. System Features

MediatorBot will be required to passively listen to discussions and intervene should an individual person be talking to much. Conversely it should also be able to detect if the conversation is slowing down and ask question to the participants, especially those that have not yet had a chance to voice their opinions. Under no circumstances should MediatorBot have an opinion on the topic or influence the participants' decision, being there only to maintain the flow of conversation at a steady level.

In addition, as people are more likely to be referenced by name[10], we plan to have MediatorBot ask people question by name. An example would be "What do you think to this John?". This is aimed to increase satisfaction with MediatorBot as it will make it seem more personal and relatable.

In addition, to reduce the amount of input that we will have to handle and the capabilities of our chosen robot platform IV-B, we have decided to focus our system to use audio and speech recognition as its main input source.

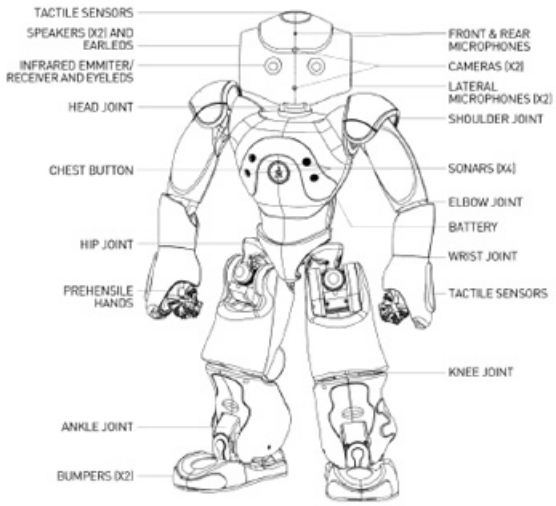

Fig. 1. Nao Schematic Diagram

To this end, MediatorBot will have to exhibit the following features:

Detect Speaker by Voice: MediatorBot will be able to detect who the speaker is from the set of known speakers using machine learning on the audio input.

Separate Multiple Speakers: MediatorBot will be able to accurately listen to the discussion even when multiple participants are speaking at the same time.

Discourage Multiple Participants from Speaking: MediatorBot will use gestures and speech to politely discourage participants from talking should they be dominating the conversation.

Locate the speaker(s): MediatorBot will be able to localise the speakers at any time within the area to turn and gesture at the correct speaker.

Balance Time Spoken between Participants: MediatorBot will be able to balance the time spoken between all participants through the use of discouraging some participants from speaking and encouraging others to speak more.

B. Robot Platform

The robot platform to be used is the autonomous, programmable robot Nao, shown in figure 1. This platform has 21 degrees-of-freedom, using under-actuated hand control. The robot features an inertial measurement unit to provide stability and positioning. Furthermore, Nao includes a multimedia system which features four microphones, two speakers and two front facing HD cameras. This robot will be used as part of the frontend of the system for interacting with the meeting participants.

In addition we require multiple microphones to provide voice localization for the different participants and also to allow for the separation of different voices speaking at the same time.

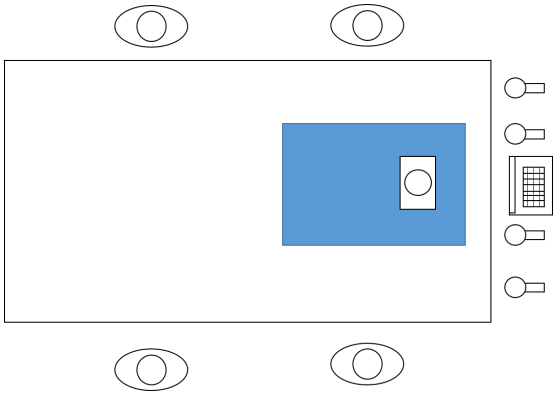

Fig. 2. Example MediatorBot Scenario. Are in blue represents the movable area.

Due to the complexity of some of the subsystems and the required external system, we have decided to move the system processing from the Nao robot onto a laptop computer. This will also allow us to use hardware acceleration and parallel computing techniques to perform more resource intensive algorithms in real time and provide are faster response to the meeting participants, a major downside to the commonly used cloud systems.

C. System Overview

The following diagram, shown in figure 3, outlines the design of the system. The system consists of 6 major areas made up of a total of 9 components.

1) *Input processing:* a module separating and cleaning up the speech signal into n number of channels, where n is the number of speakers. This is done using a Singular Value Decomposition (SVD) technique to separate the signal, using n number of microphones. This approach is irrespective of the location of the microphones, allowing for them to be hidden and increase the perceived intelligence of the robotic actor.

2) *Individual speech processing*: a speech transcription services taking the separated speech channels as input from the preprocessing module. Speech diarization is carried out to identify the speaker corresponding speech channel in order to allow for participant specific responses from the AI module. This system is aimed to run on the host laptop using a Gaussian Mixture Model (GMM) based implementation to provide real time knowledge of who is currently speaking.

3) *Natural Language Processor*: a language processing module to derive meaning from the transcribed speech. This is used to identify key words and events to be sent to the AI module. This section is aimed at being run externally on the cloud as it does not have to be instantly available and will be useful only for higher level analysis or if a single person as been talking for a significant time.

4) *AI*: a module determining how the Nao robot interacts with the participants of the conversation based on what has been said. Input is taken from the natural language processing module and used to form a decision on what relevant response is to be taken from the interaction library. This module is looking for keywords and disproportionate levels of participation from the subjects in order to trigger actions.

5) *Interaction Library*: a collection of actions and phrases for the Nao to use in order to respond to the conversation participants. These are predetermined and can be queried by the AI module when looking for a relevant response. Ideally this should have multiple responses for the same criteria so that the responses are not repeated and the whole interaction feels more natural.

6) *User Interface*: a simple application on a tablet or phone that displays information received from the host computer. For example, phrases to accompany the current actions being carried out by the Nao robot, a display of the transcribed speech that can be reviewed at the end and a graphical representation of the current amount of talking by each participant. This is important as we theorise people will be more likely to follow a robots direction if evidence is accompanied to validate its actions, to avoid suspect of bias. The display will also be used for debugging purposes, meeting analysis, and to allow us to override robots actions should it identify participants wrong or make an invalid decision.

V. EVALUATION

To validate our hypothesis we need to establish whether using a robot mediator instead of a human version to direct a conversation will lead to a more efficient and satisfactory discussion. These experiments will be conducted in a variety of round-table settings with plans to partner with the debate society, ourselves within group meetings and boardroom meetings in a professional environment as well as several staged and prompted conversations. Real world conversations with realistic constraints are more likely to give us accurate feedback on what level of intrusion by a mediator is practical and tolerated, in a range of settings and over wide user demographics. However staged prompted conversations are replicable and allow more effective comparison of the two mediator techniques.

We opted for repeated-measure design; each staged test group will need to conduct two conversations, respectively mediated by a human and robot. There will be no control group without a mediator as we are working under the assumption based off research that current human mediators do improve a conversations quality. The conversations we will be focusing on will be formal and productive by nature, with some form of outcome sought; a question is posed and a discussion ensues. The conversations will not have a inherent time limit, as one of the measures of success could be how quickly a group reaches a resolution. These same two conversations will be conducted by other groups but with different respective mediators. It would not be possible to replicate the same conversation twice within one group as the second conversation would simply be a calmer repetition of the first. Also we do not want too many conversations with the same group as they will become accustomed to each other and settle into a pattern.

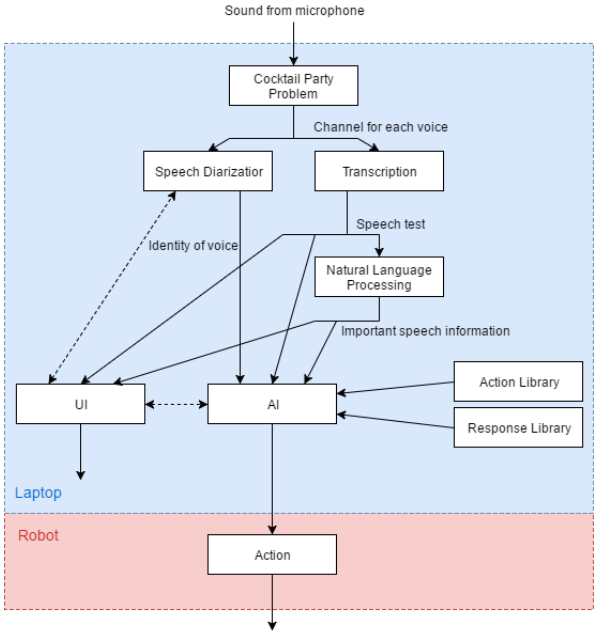

Fig. 3. Mediator System Diagram

To allow effective mediator comparison there are a number of contextual influences and user demographic differentiation that we'll need to take into consideration. Firstly it is important that the two conversations a group receive are of similar importance to them and equally controversial; for example a discussion about abortion is going to lead to a far more impassioned response than one about the best colour of a rainbow. Secondly we'll need to take into consideration the demographics of the group we are trialing with, people who are more familiar with technology are more likely to be accepting of a robotics based mediator. As such it is important we do not trial simply with Imperial students, they will not be an accurate representation of the total population's

attitude towards robotic mediators.

To determine and compare the different mediators effectiveness we need to formalise what we mean by a more successful discussion. A conversation is more efficient if there is balanced participation and input from all participants, where people are not talked over and feel free and welcome to contribute their thoughts, a pleasant tone and moderate volume is integral to this. Also if the conversation reaches an agreed conclusion quicker, with less repetition of similar discussion points.

We will determine the success of each mediation option firstly using informal feedback. This will include people's descriptions of the experience. What it was in particular they found most effective? Did they find the mediator's presence intrusive? This will be accompanied by subjective scales to represent: how well participants thought the conversation went, what level of positive impact did the mediator make, etc. The scales will be from 1-10 and will hopefully allow somewhat fairer comparison between different conversations, especially when averaged across the entire speaker group.

However it needs to be acknowledged, individual speakers will be quite biased in their feedback. For example a person who is used to getting their way, and finds the mediator prevents them from controlling the conversation is likely to have a negative perception of the mediator's role. As such the informal input will be considered on average and be primarily used to formulate the edge tolerances of intrusion that people are willing to accommodate in their conversation and overall satisfaction level. In response to this we can optimise accordingly.

We will also analyse the unbiased data recorded during the experiments to provide a reliable measurement model. This will include a transcript of the conversation, which can be evaluated for the proportion each person speaks using a simple word count, with a near equal balance being the desired output. Using natural language processing libraries it could be possible to extract the tone of the conversation, here emotions and their frequency expressed could be isolated. It could be a way to fairly analyse at least partly how pleasant an experience the conversation was, success criteria here would be a conversation with minimal angry and upset tones. The number of incidences where people are recorded talking over each other/ requiring warning to keep the conversation low and civil could be a good measure of the necessity of the mediator in the conversation. It is important that a conversation isnt just good because it happened to be a good conversation even without a mediators interference. Here we could also look at whether on average the frequency of these incidences decreased across the course of the conversation, indicating an effective mediator impact. As an extension and depending on what cloud based library services we have available to us, key word extraction could allow us to evaluate levels of repetition in the conversation.

Overall if the data collected reveals that there is no significant improvement in the balance and level tone of a conversation with a robot mediator compared to a human mediator and there is no informal preference for the robotics system then our hypothesis will have been falsified and deemed incorrect.

VI. CONCLUSION

In this design report we have discussed the use of a robot as an unbiased mediator to make business meetings more efficient. The design outlined consists of a small humanoid robot, the NAO, and a user interface displayed on a laptop to interacting with participants depending on the analysis of the conversation recorded by a microphone.

Should the evaluation of our design prove our hypotheses, MediatorBot could greatly improve the efficiency of meetings, saving time and money and accelerating the success of businesses.

REFERENCES

[1] D. J. Leach, S. G. Rogelberg, P. B. Warr, and J. L. Burnfield, "Perceived meeting effectiveness: The role of design characteristics," *Journal of Business and Psychology*, vol. 24, no. 1, pp. 65–76, 2009.

[2] K. D. Benne and P. Sheats, "Functional roles of group members," *Journal of Social Issues*, vol. 4, no. 2, pp. 41–49, 1948.

[3] S. Kaner, L. Lind, C. Toldi, S. Fisk, D. Berger, and M. Doyle, *Facilitator's Guide to Participatory Decision-Making*. Wiley, 2007.

[4] Y. Matsuyama, I. Akiba, S. Fujie, and T. Kobayashi, "Four-participant group conversation: A facilitation robot controlling engagement density as the fourth participant," *Computer Speech & Language*, vol. 33, no. 1, pp. 1 – 24, 2015.

[5] R. Schwarz, *The skilled facilitator: A comprehensive resource for consultants, facilitators, managers, trainers, and coaches*. John Wiley & Sons, 2002.

[6] T. Kim, A. Chang, L. Holland, and A. S. Pentland, "Meeting mediator: Enhancing group collaborationusing sociometric feedback," in *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, (New York, NY, USA), pp. 457–466, ACM, 2008.

[7] Y. Zhang, M. Bern, J. Liu, K. Partridge, B. Begole, B. Moore, J. Reich, and K. Kishimoto, *Ubiquitous Meeting Facilitator with Playful Real-Time User Interface*, pp. 3–11. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[8] G. Tur, A. Stolcke, L. Voss, S. Peters, D. Hakkani-Tur, J. Dowding, B. Favre, R. Fernandez, M. Frampton, M. Frandsen, C. Frederickson, M. Graciarena, D. Kintzing, K. Leveque, S. Mason, J. Niekrasz, M. Purver, K. Riedhammer, E. Shriberg, J. Tien, D. Vergyri, and F. Yang, "The calo meeting assistant system," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, pp. 1601–1611, Aug 2010.

[9] Microsoft, "Meeting recognition and understanding." `https://www.microsoft.com/en-us/research/project/meeting-recognition-and-understanding`, 2016. Online; Accessed: 2016-11-02.

[10] D. P. Carmody and M. Lewis, "Brain activation when hearing one's own and others' names," *Brain research*, vol. 1116, no. 1, pp. 153–158, 2006.