

# A user study of a humanoid robot as a social mediator for two-person conversations

Tahir, Yasir; Dauwels, Justin; Thalmann, Daniel; Thalmann, Nadia Magnenat

2018

Tahir, Y., Dauwels, J., Thalmann, D., & Thalmann, N. M. (2018). A user study of a humanoid robot as a social mediator for two-person conversations. *International Journal of Social Robotics*. doi:10.1007/s12369-018-0478-3

<https://hdl.handle.net/10356/141981>

<https://doi.org/10.1007/s12369-018-0478-3>

---

© 2018 Springer Science+Business Media B.V., part of Springer Nature. This is a post-peer-review, pre-copyedit version of an article published in *International Journal of Social Robotics*. The final authenticated version is available online at:  
<http://dx.doi.org/10.1007/s12369-018-0478-3>

*Downloaded on 13 Mar 2024 18:13:36 SGT*

# A User Study of a Humanoid Robot as a Social Mediator for Two-Person Conversations

Yasir Tahir · Justin Dauwels · Daniel Thalmann · Nadia Thalmann

Received: date / Accepted: date

**Abstract** In this work we have enhanced the perception of a humanoid robot by integrating it with a social state estimation system. We present a user study of the humanoid Nao robot as a social mediator, comprising two sets of experiments. In the first sets of experiments, the participants rate their understanding of feedback messages delivered via the Nao robot. They also assess two modalities to deliver the feedback: audio only and audio combined with gestures. In almost all cases there is an improvement of 10% or more when audio and gesture modalities are combined to deliver feedback messages. For the second sets of experiments the sociofeedback system was integrated with the Nao robot. The participants engage in two-person scenario-based conversations while the Nao robot acts as a mediator. The sociofeedback system analyzes the conversations and provides feedback via Nao. Subsequently, the participants assess the received sociofeedback with respect to various aspects, including its content, appropriateness, and timing. Participants also evaluate their overall perception of Nao as social mediator via the Godspeed questionnaire. The results indicate that the social feedback system is able to detect the social sce-

nario with 93.8% accuracy and that Nao can be effectively used to provide sociofeedback in discussions. The results of this paper pave the way to natural human-robot interactions for social mediators in multi-party dialog systems.

**Keywords** Sociometrics · Dialog · Audio-Visual · Human Behavior.

## 1 Introduction

One of the key objectives of research and development in robotics is to design various robots that can assist humans in everyday domestic environments. Nowadays, robots are increasingly being viewed as social entities to be integrated in our daily lives. Socially interactive robots are used to communicate, express, and perceive emotions, maintain social relationships, interpret natural cues, and develop social competencies [1, 2]. Prominent application scenarios for such robots are manifold, and span from shopping robots [3] and tour guides [4] to home assistance and care [5, 6], etc.

With increasing demand for robots for domestic environments, research on human-robot interaction (HRI) has gained more importance. In order to enhance human-robot interaction, the need for integration of social intelligence in such robots has become a necessity [7–9]. Socially intelligent robots should effectively engage with humans and maintain a natural interaction with them over extended periods of time.

Understanding of human behavior is a necessary requirement for allowing a robot to behave in a socially intelligent manner [10]. If a robot can understand the behavior of humans with whom it is interacting, then it can respond accordingly. HRI in multi-party dialogs [11]

---

Yasir Tahir  
Institute for Media Innovation, Nanyang Technological University.  
E-mail: yasir001@e.ntu.edu.sg

Justin Dauwels  
School for Electronics and Electrical Engineering, Nanyang Technological University.

Daniel Thalmann  
Institute for Media Innovation, Nanyang Technological University.

Nadia Thalmann  
Institute for Media Innovation, Nanyang Technological University.

can be greatly improved if the robots are able to interpret the human behavior to some extent. Human behavior involves various patterns of actions and activities, attitudes, affective states, social signals, semantic descriptions and, contextual properties [12]. A promising approach for human behavior understanding is to apply pattern recognition and automatically deduce various aspects of human behavior from different kinds of recordings and measurements, e.g., audio and video recordings [13].

In [14], we presented a novel approach towards comprehensive real-time analysis of speech mannerism and social behavior. We performed non-verbal speech analysis to analyze human behavior. Non-verbal speech metrics are a direct manifestation of human behavior, and play a vital role for the meetings to be pleasant, productive, and efficient [15]. By considering these low-level speech metrics, we quantified speech mannerism and sociometrics including interest, agreement, and dominance of the speakers. We collected a diverse speech corpus of two-person face-to-face conversations; it allowed us to train machine learning algorithms for reliable 5-level classification of the sociometrics with speech metrics as input features. The classifier is able to detect social states of participants with accuracy of 84–86%. The combined metrics for speech mannerism and social behavior provided a clear picture of human behavior in dialogs. In this paper, we investigate the scenario where the Nao robot communicates this information to the speakers and acts as a “social mediator”.

In [16], we conducted a preliminary user study to investigate how sociofeedback could be provided via a humanoid robot (Nao). It is widely accepted that the combination of modalities and capabilities improves human-robot interaction. In our preliminary study, we investigated a variety of modalities. We provided users with sociofeedback in open-loop conditions. Specifically, the participants of the survey needed to assess basic feedback messages delivered by Nao, without actually participating in a conversation. The participants were then asked to assess sociofeedback messages delivered only via audio and also by a combination of audio and gestures. The user study confirmed the hypothesis that combining the two modalities of audio and gestures clearly helps the participants to identify the sociofeedback messages.

In this paper, we extend our work from the open-loop to closed-loop scenario. In the current study, the participants have a conversation, and the Nao robot provides feedback afterwards. This feedback is derived from the speech mannerisms and sociometrics computed by the machine learning algorithms proposed in our earlier work [14]. In other words, the Nao robot serves in

this setting as a social mediator, and we are interested to study how the participants react to and evaluate this approach to providing feedback. This paper presents the following contributions and novelties:

- We integrate a real-time sociofeedback system that analyzes nonverbal speech metrics to assess the social states of participants in a two-person conversation with a humanoid robot (Nao). The robot uses this information and provides appropriate feedback in real-time. Currently, we limit ourselves to four social states, namely normal, uninterested, overly talkative, and aggressive.
- We conducted a user study with 20 participants (17 males, 3 females). Each participant received sociofeedback via Nao for all the social states. Participants were then asked to evaluate several aspects of sociofeedback e.g., whether they *agree* with the feedback, or whether they *like* the feedback, whether they feel they received the feedback *timely*.
- We also investigated the overall experience of users about Nao. The participants were asked to rate the anthropomorphism, animacy, likability, perceived intelligence and perceived safety by means of a Godspeed questionnaire [17].

In summary, we made the following observations in our experiments. From our first experiment [16], we learned that the participants could clearly understand the feedback messages delivered by Nao using gestures along with audio. The ratings on the Godspeed questionnaire were also significantly higher for the case when Nao used audio and gestures to deliver feedback message as compared to only using audio. In the second experiment we observed that the participants seemed to like Nao in the role of a social mediator and rated it very high on the Godspeed questionnaire. We also established that the sociofeedback system [14] can provide reliable feedback in real conversational scenarios with an overall accuracy of 93.8%.

In order to avoid background noise, we performed these experiments in a meeting room scenario where the participants wore lapel microphones. As a consequence, the recorded audio signals are of high quality, and we can infer the social states reliably. In the current study, we apply the sociofeedback system that we designed in earlier work. Since that system only infers a limited number of social indicators (level of interest, dominance, and agreement), we concentrate only on a limited number of social states in the present study. Our objective in the long term is to design a module for inferring social states for applications in robotics, and the experiments discussed in this paper are initial steps towards achieving this objective. Here we imple-

mented our system on the humanoid Nao robot, due to its availability and ability to perform gestures and generate speech. However, the sociofeedback system is not limited to the Nao robot and can be interfaced with other robot platforms as well as virtual characters. Similarly, we apply the Godspeed questionnaire [17] in this study, however, it is noteworthy that this questionnaire is not specific to the Nao robot, but can be applied to assess any robotics platform. The technical aspects of this work are also not dependent on the Nao platform as we utilize the sociofeedback system [14] for acquiring and processing the speech signals and for identifying the social states of the participants.

The paper is structured as follows. In Section 2, we review related work. In Section 3, we present a brief overview of the sociofeedback system that infers the level of interest, dominance and agreement from speech recordings. In Section 4, we briefly introduce the two sets of experiments that we have conducted. In section 5, we present results for our first set of experiments, where we determine whether the participants can identify feedback messages delivered by Nao. In section 6, elaborate on the results for our second set of experiments, where we interfaced Nao with the sociofeedback system. In Section 7, we offer concluding remarks and suggest topics for future research.

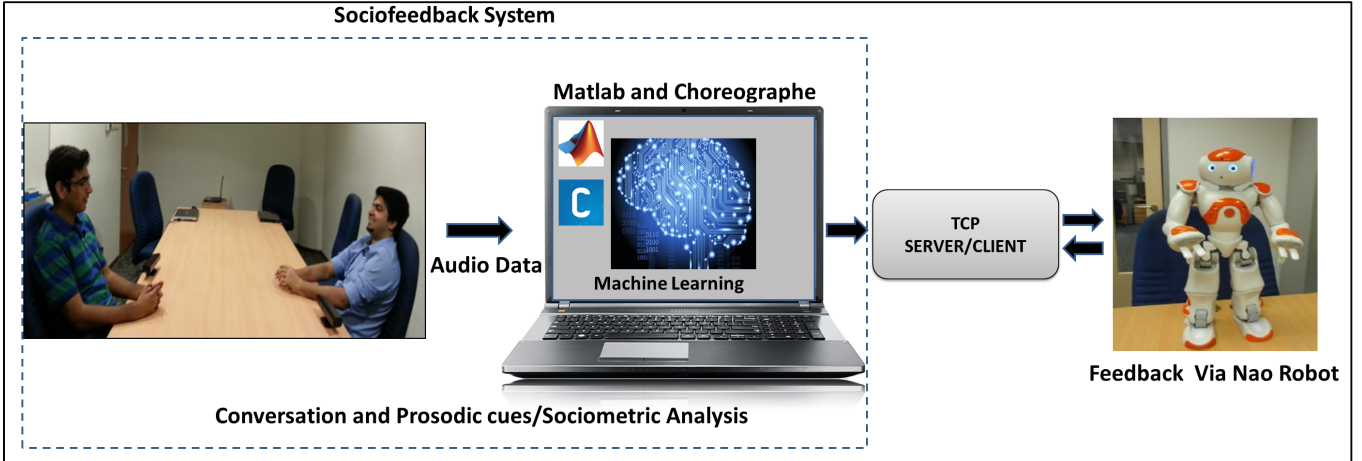
## 2 Related Work

In this section, we briefly discuss related work on socially aware robotic systems, their applications, and relevant user studies to assess human-robot interaction. In the recent past, many social robots have been designed for real world interactions, e.g., Kismet [18], Mel [19], Pearl [20], Robovie [21], Robota [22], and Paro [23]. Nowadays, social robots are successfully helping children in their social, emotional, and communication deficits. They create interesting, appealing, and meaningful interplay situations that compel children to interact with them. One of the emerging applications of social robotics is the therapy of children with autism [24–26]. The roles and benefits of socially aware robots for therapy of children with autism are reviewed in [27]. Similarly, social robots are actively being deployed in nursing homes for assistance of the elderly. Those studies typically investigate what different social functions such robots can play in the living environment of the elderly, as well as how social functions can facilitate actual usage of social robots [23].

Apart from that, many application centric social robots are being deployed in domestic environments where the goal is to interact with humans as naturally as possible. The Human-Computer Interaction Institute

(HCII) at Carnegie Mellon University (CMU) has developed an advisory robot that traces people’s mental mode from a robot’s physical attributes [28]. Similarly, the iCat Research Platform is a research platform created by Philips Electronics for studying human-robot interaction. The robot itself consists of a catlike robot face with two mechanical eyes, eyebrows, eyelids, and lips, all attached to a limbless body. iCAT has been deployed in [29] to investigate dynamic multi-party social interaction with a robot agent. CALO-meeting assistant is an automatic agent that assists meeting participants, and is part of the larger CALO [30] effort to build a Cognitive Assistant that Learns and Organizes. CALO meeting assistant [31] provides for distributed meeting capture, annotation, automatic transcription, and semantic analysis of multiparty meetings. As a last example, Furhat is a robotic head that combines state-of-the-art facial animation with physical embodiment in order to facilitate multi-party dialogues with robots [32].

Many user studies have been conducted to assess how humans perceive robots in their specific roles. Such studies rate the human-robot interaction with respect to likability, perceived safety, anthropomorphism, animacy, etc. For example, it was investigated in [33] how humans perceive affect from robot motion, and they found that many participants engaged in seemingly emotional and unexpected ways with a very simple and almost purely abstract robot. It was shown in [34] that humans perceive different affects by observing different motions of the robot. The curvature and acceleration of robot motion were varied and their positive or negative affect on the participants were observed. The results indicate that the information for valence is at least partly carried by a linear interaction between curvature and acceleration. Similarly, in [35] studies have been carried out to see if humans can identify emotions expressed by a humanoid robot using gestures. The results show that it is possible to interpret key poses generated by the Affect Space. This suggests that the approach can be used to enrich, at a low cost, the expressiveness of humanoid robots. In [36] Nao narrated a three-minute story to a group of participants. The study investigated the effect of gazing and gestures on the persuasion of the robot, and provides evidence that gazing can significantly improve persuasion, however, incorporating gestures showed no significant difference in persuasion. In [37], experiments were carried out to understand whether a robot can effectively modify its speech according to the speaker’s behavior. This study offers a model for dealing with the emotions in the voice of the user in an AI system. Low-level cues computed from the speech signals determine characteristics of the ex-



**Fig. 1** The system records audio data, and next computes several conversational and prosodic features. From those features, it determines the levels of interest, agreement, and dominance via classifiers. Feedback messages are determined from these three social indicators and from prosodic features. All these computations are performed in Matlab. The feedback messages are communicated from the computer to Nao via the TCP/IP framework. The Nao robot provides feedback by an audio message supported by gestures. The gestures are programmed in Choreographe.

pressed emotions, such as the emotion type (e.g., happiness and sadness), its valence, and its activation.

By contrast, our objective is to facilitate multi-party dialogs by introducing Nao as a social mediator, which can assess social state of participants, in real-time, and provide valuable feedback without having to provide any service or engage participants in any context-based conversation. To achieve this, we conducted a study to investigate, in detail, different aspects of human-robot interaction when Nao provides real-time sociofeedback to participants. To the best of our knowledge, no such study has been conducted yet.

### 3 The Sociofeedback System Overview

In Fig. 1, we depict a diagram of the robotic system considered in this paper. The system consists of the sociofeedback module (developed in [14]) and the Nao robot, where Nao delivers the messages generated by the sociofeedback module to the speakers participating in dialogues. In this section, we will explain the sociofeedback system [14]. This system is able to infer the levels of interest, dominance, and agreement with 85%, 86% and 82% accuracy respectively (see Fig. 2). In the following subsections, we first explain the hardware setup for audio recording of conversations. Next, we briefly describe the extraction of nonverbal speech cues. Then, we explain how we infer social states from those cues. Finally, we explain how the sociofeedback interfaces with the Nao robot to provide real-time sociofeedback.

#### 3.1 Sensing and Recording

We adopted easy-to-use portable equipment for recording conversations; it consisted of lapel microphones for each of the two speakers and an audio H4N recorder that allowed multiple microphones to be interfaced with the laptop. The audio data was recorded in brief consecutive segments as a 2-channel audio .wav file.

#### 3.2 Extraction of Non-Verbal Cues

We considered two types of low-level speech metrics: conversational and prosody related cues. The conversational cues account for *who* is speaking, *when* and *how* much, while the prosodic cues quantify *how* people talk during their conversations. We computed the following conversational cues: the number of natural turns, speaking percentage, mutual silence percentage, turn duration, natural interjections, speaking interjections, interruptions, failed interruptions, speaking rate, and response time [14].

We considered the following prosodic cues: amplitude, larynx frequency (F0), formants (F1, F2, F3), and mel-frequency cepstral coefficients (MFCCs). These cues are extracted from 30ms segments at a fixed interval of 10ms; they tend to fluctuate rapidly in time. Therefore, we compute various statistics of those cues over a time period of several seconds, including minimum, maximum, mean, and entropy, in order to infer speaking mannerism.

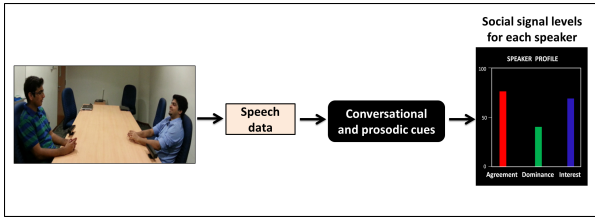


Fig. 2 Diagram of the sociofeedback system designed in [14].

### 3.3 Social State Estimation

In our earlier work [14], we collected an audio corpus of 150 conversations where the subjects were students of Nanyang Technological University (NTU). The total number of individuals that participated in the corpus was 22, of which 17 were males, and 5 were females. The age of the students varied from 18 to 30. The topics of conversations ranged from discussion of assignments, projects of students, to social and political views. In some of the dialogs, there were problematic situations such as conflicts and disagreements, periods of boredom, aggressive behavior, or poorly delivered speech (e.g., low volume or fast pace).

Each recording in the corpus was annotated by multiple people (“judges”), each assessing a subset of the corpus. For each recording in the corpus, the judges completed a questionnaire related to speaking mannerisms and behavioral aspects of each participant. For example, if a participant seemed bored to the annotator, the latter would assess the interest level as “low”; in contrast, if the participant seemed excited, the annotator would quantify the interest level as “high”.

Low-level speech cues and conversational features were extracted from each recording. It is crucial to select appropriate speech features for training a machine learning model. We applied two feature selection algorithms Information Gain (IG) and correlation based feature selection (CFS) [38,39], to determine the most relevant features for inferring each of the three sociometrics i.e. interest, dominance and agreement.

After feature selection the speech features were used to train machine learning algorithms. The (rounded) average score provided by the judges served as labels for supervised learning. We considered four kinds of multi-class classifiers for inferring the social state of the participants: K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), Naive Bayes, and Support Vector Machine (SVM). Table 1 shows the classification results achieved by the aforementioned algorithms, and Table 2 shows the detection accuracies for high/low volume and fast/slow speech rate speech mannerisms. Speaking mannerism are quantitatively assessed by low-level speech, including volume and speech rate. The

Table 1 The classification results achieved for each sociometric using various machine learning algorithms.

Sociometrics	SVM	ANN	KNN	Naive
Agreement	83%	76%	82%	78%
Dominance	86%	80%	78%	81%
Interest	82%	78%	79%	74%

Table 2 Detection accuracies for the speech mannerisms of high/low volume and fast/slow speech rate.

Speech Mannerism	Audio Feature	Detection (%)
Speaking loudly/quietly	Volume	90%
Speaking too fast/too slow	Speech Rate	84%

SVM algorithm performed better than other classifiers for all the three sociometrics.

We performed these calculations in Matlab on a 2GHz dual-core processor with 2GB RAM. It took approximately 3-5 seconds to perform speech detection and compute speech cues from 1 min dialogs, and to perform multi-class classification, yielding the levels of interest, agreement, and dominance. Therefore, on that computer platform, the total time required for inferring those social indicators from a 1 min dialog is about 3-5 seconds, allowing us to perform such analysis in real-time settings with limited delay.

In this paper, we conduct a user study of this sociofeedback system [14], by integrating it with a humanoid Nao robot. We deploy SVM models trained on a corpus of 150 conversations [14] from our earlier work. The participants engage in 1 minute conversations and receive feedback via Nao robot at the end of the dialog. They then evaluate the sociofeedback system on various criteria. The audio data is acquired using lapel microphones, it is processed on the laptop, and the sociometrics are computed. The social behavior is quantified by the level of interest, agreement, and dominance. Together, they provide a comprehensive picture of the social state of participants in these dialogs. The inferred social state is then used to generate feedback via Nao robot.

### 3.4 Feedback via Nao Robot

The social states of the speakers listed in in Table 4 are derived from the social indicators of interest, dominance, and agreement as explained in Table 7. These social states are computed by a Matlab script, as explained in the previous section, and are transmitted to the Nao robot by the TCP/IP server-client framework. More precisely we integrate Nao into this sys-

tem by transmitting the output of the Matlab script to Nao through the TCP/IP server-client framework. More precisely, once the Matlab script determines the social state, it sends a feedback message to Nao via TCP/IP, and Nao in turn delivers the message to the speaker(s) via speech supported by gestures.

The Nao robot has 25 degrees of freedom, since it is equipped with numerous sensors and actuators, including inertial sensors, infrared and sonar receivers, coupled with its axes. This multitude of sensors and actuators provide the robot with high level of stability and fluidity in its movements. However, in our experiments we only generated very basic movements to simulate gestures. In addition, we also utilized the speech synthesis module of the Nao robot to generate audio messages. In some of our experiments, Nao delivered the audio messages without gestures, in other experiments Nao robot made gestures during the audio messages. The time taken by Nao to deliver the audio message along with with gestures was approximately 3 to 4 seconds. Table 4 provides an overview of the feedback messages considered in this study. We chose these particular scenarios, since the sociofeedback system was trained on a corpus of dyadic conversation with similar scenarios [14].

## 4 Experiments

In this section, we explain the two sets of experiments that we conducted. In the first set of experiments (see Section 4.1), we investigate whether the participants can understand the feedback messages delivered by the Nao robot. We also explore different ways to deliver the feedback messages, viz., only by audio or by audio combined with gestures. In the second set of experiments (see Section 4.2), we investigate the role of the Nao robot as a social mediator. In this setting, the Nao robot provides feedback to the participants after brief conversations. The first experiment is a pre-requisite to the second one. In the first experiment we obtain user feedback on whether our designed gestures and verbal messages can be understood. The second experiment includes more complexity as explained in Section 3, the participants engage in scenario-based conversations. The latter are then analyzed by the sociofeedback module [14], specifically, SVM models trained on an audio corpus of 150 annotated conversations are deployed to predict the levels of interest, dominance, and agreement for each of the participants from their speech features. The social state of the speaker is estimated from the three social indicators (interest, dominance, and agreement), and the corresponding feedback is delivered via the Nao robot. Table 3 lists the user studies that we

have conducted along with the objectives of each user study.

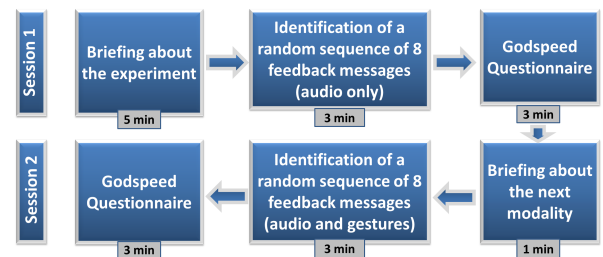
**Table 3** List of experiments conducted and their objectives.

Experiments	Objectives
Experiment 1: Identification of Feedback Messages	1-To determine the accuracy with which the participant can identify feedback messages delivered by Nao.
	2-To compare audio and gesture modalities for feedback delivery , assessed by Godspeed questionnaires.
Experiment 2: Integration with the Sociofeedback System	1-To determine the accuracy with which the sociofeedback system can analyze and generate feedback messages for real conversations.
	2-To investigate how the participants assess the Nao robot as a social mediator, by means of Godspeed questionnaires.

In our experiments we obtained feedback from the participants about the Nao robot by means of Godspeed questionnaires [17]. In the Godspeed questionnaire the participants rated their perception of the robot on different criteria. Including anthropomorphism (similarity to human form), animacy (life likeness), likeability (personal likeness of the participant), perceived intelligence, and perceived safety of the robot.

### 4.1 Experiment 1: Identification of Feedback Messages

There were 20 (16 males and 4 females) participants in this first set of experiments with a mean age of 25 and SD of 2.42. All participants are NTU students. As the medium of instruction at NTU is English, all participants could easily understand the feedback messages delivered by Nao robot. The experiments were conducted in a meeting room similar to the one shown in Fig. 1.



**Fig. 3** Different components of the experimental procedure. The experiments last about 20 minutes, with estimated duration of each component as indicated.



**Table 5** Percentage of correctly identified feedback messages. Results are shown for each of the feedback messages, delivered by audio only messages and by a combination of audio and gestures.

Modality	Too silent	Too loud	Aggressive	Overly talkative	Uninterested
Audio	84.2%	68.4%	89.4%	89.4%	78.9%
Combined	84.2%	100%	94.7%	100%	100%

Each experiment in the first set lasted about 20 minutes, and comprised of two sessions (see Fig. 3). First the participants were asked to identify a random sequence of eight messages that were delivered by audio only (without gestures), next the same is repeated for messages delivered by audio and supported by gestures. After each message, there was a brief break in which the participants selected the feedback message that they believed the Nao robot had just delivered. The participants were given the possible answers, and they had to choose one of them. After each of the two sequences of eight messages, the participants were asked to complete a Godspeed questionnaire about their experience with the Nao robot.

In this experiment, the participants were not asked to be a part of an active conversation; instead they were briefed about the context, and were then asked to judge the sociofeedback delivered by the Nao robot (feedback messages illustrated in Table 4). The participants were not informed about the correct answers after each session, in order to minimize the learning effect.

With these experiments, we aimed at testing the hypothesis that feedback messages can be identified more accurately when Nao uses gestures along with audio as compared to only audio feedback. We also hypothesized that the feedback delivered by both audio and gestures will be rated higher on Godspeed questionnaire criteria, as compared to feedback provided by only audio messages.

In our questionnaire (see Table 6), there were two questions associated with each of the five measures of Godspeed questionnaire. Our questionnaire was a subset of the original Godspeed questionnaire, as we wanted to keep the experiments short.

#### 4.1.1 Results for Feedback Identification

Our results are summarized in Table 5, showing how often (percentage) each of the feedback messages were correctly classified in each of the sessions. It can be seen from Table 5 that most of the feedback messages seem to be perfectly understandable when the audio messages are combined with gestures. There is room for improvement for the “Too silent” scenario. It is also







clear from Table 5 that combining audio messages with gestures helps to improve the clarity of the feedback messages, as compared to audio messages only. To verify whether this improvement is statistically significant, we applied a repeated measures single-factor ANOVA statistical test to the responses of the 20 participants. The p-value associated with audio only vs. combined audio and gestures equals 0.027, which is clearly below 0.05, thus the corresponding improvement in accuracy is indeed statistically significant. These statistics indicate that sociofeedback is easier to identify when delivered through both audio and gestures. Also the results show that audio plays a more vital role in the delivery of the feedback messages while gestures help in improving the clarity of the message.

#### 4.1.2 Results for the Godspeed Questionnaire

Our results are summarized in Table 6, showing the average scores for both conditions and the corresponding p-values of repeated measures single-factor ANOVA test. It can be seen that for each of the five measures (except perceived safety), at least one of the two questions is having a significant change in its value. Specifically, the Godspeed scores are higher for feedback that includes both audio and gestures. The score for likeability is the highest, and the change in value for friendliness is significant. In other words, the participants seem to like the Nao robot, and by including gestures, the robot is perceived as even more friendly. Anthropomorphism also has good ratings, and the increase in values by adding gestures is significant for both questions. Moreover, the interactivity of the robot increases significantly when gestures are included. Likewise, the participants perceive the robot as more knowledgeable when it uses gestures, but the value does not change significantly for the intelligence shown by the robot. The low perceived safety values suggest that the participants were calm and quiescent in the presence of the robot, since the minimum and maximum value correspond to calmness and agitation respectively. Interestingly, when gestures are added, the participant perceived the robot’s behavior as slightly more safe (albeit a small change).



**Table 4** Sociofeedback delivered by the Nao robot: gestures (left) and speech (right).

Gestures	Description
	<b>Normal:</b> “Good, carry on.” Nao provides this feedback when a smooth conversation is going on.
	<b>Uninterested:</b> “You both seem uninterested.” Nao will invite the speakers to contribute more to the discussion, when both speakers have not been speaking for a period of time.
	<b>Overly talkative:</b> “You are talking a lot”. Nao will ask the speaker to slow down when he/she is speaking too much.
	<b>Aggressive:</b> “Please calm down”. Nao will ask the speaker to calm down if he/she is being too aggressive.
	<b>Too silent:</b> “I am sorry, but I cannot hear you”. When one or both of the speakers are speaking too softly, Nao will ask them to increase their volume.
	<b>Too loud:</b> “Please lower your volume”. When the speakers are speaking too loudly, Nao will give feedback about the noise.

**Table 6** Average values of Godspeed questionnaire(5-likert scale).

Characteristics	P-values	Average (audio)	Average (combined)
<b>Anthropomorphism</b> Machine/human like	<b>0.017</b>	2.94	3.52
Moving rigidly/elegantly	<b>0.002</b>	2.36	3.26
<b>Animacy</b> Mechanical/organic	0.129	2.89	3.26
Inert/interactive	<b>0.046</b>	3.15	3.63
<b>Likability</b> Dislike/like	0.110	4.26	4.63
Unfriendly/friendly	<b>0.008</b>	3.73	4.26
<b>Perceived Intelligence</b> Ignorant/knowledgeable	<b>0.009</b>	3.31	3.63
Unintelligent/intelligent	0.186	3.42	3.57
<b>Perceived Safety</b> Calm/agitated	0.741	2.36	2.26
Quiescent/surprised	1	2.84	2.84

#### 4.2 Experiment 2: Integration with the Sociofeedback System

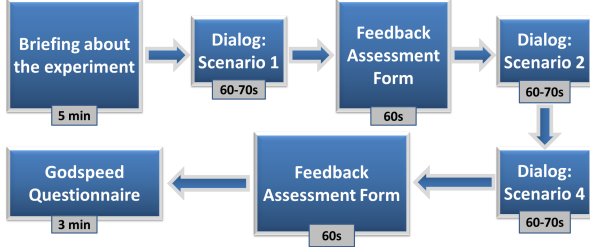
There were 20 (17 males and 3 females) participants in the second set of experiments with a mean age of 23 and standard deviation of 2.42. The total duration for each experiment session was around 20 minutes. The aim of this second set of experiments is to investigate whether Nao can interact as a social mediator in a two-person dialog. We invited participants to have a scenario-based conversation. In each scenario, the participants were asked to behave according to four scenarios: “normal”, “uninterested”, “overly talkative”, and “aggressive”, corresponding to the first four situations listed in Table 2. The bottom two scenarios in that Table are less interesting as social states, and hence are not considered in this experiment. In order to facilitate the scenario-based conversations, we asked the participants to follow scripted conversations. From our earlier experiments, we learned that it is difficult for participants to enact a scenario if both speakers are invited subjects. Therefore, in this user study we changed our approach such that one of the two speakers was an invited participant while the other speaker was appointed by us to serve as facilitator of the conversations. Each conversation lasted about 60 to 70s, and was analyzed in real-time by the sociofeedback system described in Section 3 (see also Fig. 1).

The experiment was conducted as follows (see Figure 4):

- First, we setup the recording system properly.

**Table 8** Percentage of correctly delivered feedback messages.

Normal	Uninterested	Overly talkative	Aggressive	Overall
100%	90%	85%	100%	93.8%

**Fig. 4** Different components of the experimental procedure. The experiments last about 20 minutes, with estimated duration of each component as indicated.

- The two speakers sat about 1.5m apart so that each microphone only recorded the voice of the respective speaker, and there was no interference from the other speaker.
- We attached the lapel microphones to the speakers in proper manner, in order to obtain a high-quality recordings.
- The participant and the facilitator had scenario based conversations. Each conversation was about one minute in duration.
- Nao robot gave feedback after each conversation, depending on the scenario.
- The participant filled a questionnaire after each conversation, in order to rate the feedback delivered by the robot.
- At the end of the experiment, the participant completed the Godspeed questionnaire in order to rate the Nao robot in the role of social mediator.

#### 4.2.1 Accuracy of Sociofeedback System

The participants were asked to act according to the first four scenarios listed in Table 2. If the feedback message delivered by Nao is in accordance with the enacted scenario, it is considered accurate. In Table 7 we present

**Table 7** Relationship between the social scenarios and the social indicators of interest, dominance and agreement.

Scenario	Interest	Dominance	Agreement
Normal	Medium	Medium	High
Uninterested	Low		
Overly talkative	High (Low for the other speaker)		
Aggressive	High	High	Low

the relationship between the social scenarios and the values of interest, dominance and agreement. In Table 8 we list the accuracy of the feedback for each scenario and also present the overall accuracy of the system. In Table 9 we show the confusion matrix for these scenarios.

**Table 9** Confusion matrix showing the classification results of first four scenarios. The feedback generated in these scenarios used interest, dominance and agreement sociometrics predicted by means of an SVM classifier.

	Normal	Uninterested	Overly talkative	Aggressive
Normal	20	0	0	0
Uninterested	0	18	0	2
Overly talkative	1	0	17	2
Aggressive	0	0	0	20

As seen from Table 8 the overall accuracy of the feedback messages is 93.8%. In the cases of “Normal” and “Aggressive” scenarios all the conversations generated correct feedback but there were mistakes in other scenarios. False detections can occur when the participants do not strictly follow the scenario. We also asked the participants whether they agreed with the provided feedback, resulting in a average score average rating of 4.5 on a scale of 5 (see Table 11). This shows that the participants mostly agreed that the feedback provided by the Nao robot was appropriate for the scenario.

#### 4.2.2 Assessment of Nao as Social Mediator

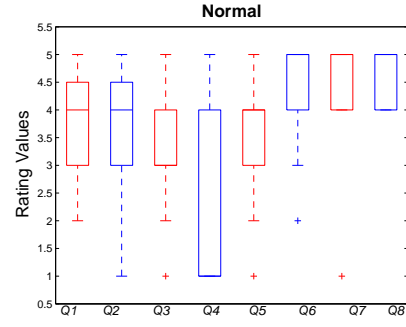
At the end of each conversation, the participants were asked to complete an assessment about the received feedback message. The questions concern different aspects of the feedback, including the content of feedback, likability, and timing (see Table 10). At the end of all the conversations, the participant rated his/her experience of Nao as social mediator via a Godspeed questionnaire. The purpose was to obtain the user opinion about the robot in the role of a social mediator. Table 12 shows the average ratings for each of the Godspeed criteria. In order to keep the assessments consistent, we adopted a 5-likert scale for both questionnaires.

**Table 10** Questions of the assessment form.

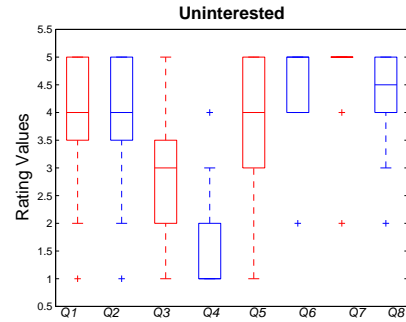
	Question
<b>Q1</b>	Did you notice when the sociofeedback system was addressing you?
<b>Q2</b>	Did you notice when the sociofeedback system was addressing others?
<b>Q3</b>	Was the timing of sociofeedback appropriate?
<b>Q4</b>	Did the sociofeedback system interrupt the conversation?
<b>Q5</b>	Was the interaction natural?
<b>Q6</b>	Did you understand the message given by the sociofeedback?
<b>Q7</b>	Do you agree with the given feedback?
<b>Q8</b>	Did you enjoy using the sociofeedback system?

Fig. 5 displays the eight ratings for each of the feedback messages. As can be seen from these figures, the ratings are mostly high. The average ratings for each question (Q1 – Q8) can be seen in Table 11. Q1 and Q2 asked the participants if they could tell when Nao was addressing them or the other speaker. The high values for all the cases implies that the participants were able to distinguish among feedback messages meant for them and the other speaker. In Q3, we asked participants about the timing of the feedback. Although most participants stated that Nao gave feedback timely, there is still room for improvement. The ratings of Q4 suggests that participants at times felt that they were interrupted by Nao. The timing can be improved by waiting for the speaker to stop his/her sentence or by getting the attention of the speaker using some gesture, before delivering the feedback message. Furthermore, the high ratings for Q5 and Q6 suggest that the interaction between Nao and the participants was fairly natural and Nao spoke with clarity. In Q7 we asked whether the participants agreed with the feedback message. The rating for this question is close to 5, indicating that participants agree with the feedback. Similarly, high ratings for Q8 confirm that participants like the feedback from Nao. Each column shows the average ratings for different scenarios.

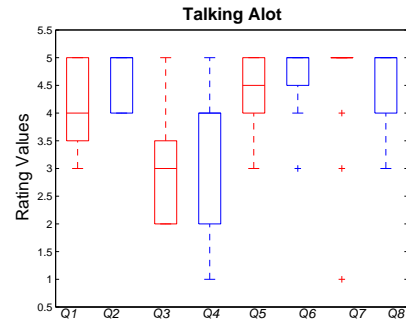
The scores for likeability are the highest. In other words, the participants seemed to like Nao, and perceived it as friendly. Anthropomorphism also has good ratings. The robot is rated strongly human-like but



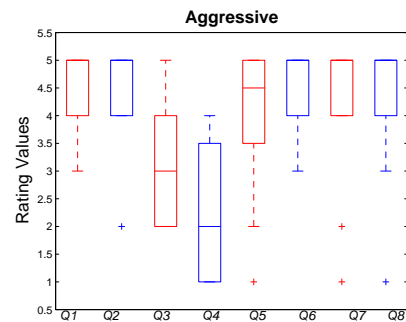
(a) Ratings for “Normal” scenario.



(b) Ratings for “Uninterested” scenario.



(c) Ratings for “Overly talkative” scenario.



(d) Ratings for “Aggressive” scenario.

**Fig. 5** Box plots of participant’s ratings for “Normal”, “Uninterested”, “Overly talkative”, and “Aggressive” scenarios.

the motions of the robot can be improved to make it more elegant. The animacy of Nao is also rated high by the participants, consequently, Nao was considered as highly interactive. Likewise, the participants perceived the robot as knowledgeable and intelligent. However, Nao received moderate ratings for its perceived safety, suggesting there is a room for improvement to make the participants more comfortable in the presence of Nao. Perceived safety is related to the size of the robot. Nao is a small robot (2 feet); when people interact with Nao while they are standing, the safety value is usually high [40]. In our case, Nao is seated very close to the participants (see Fig. 1), which may explain why the safety value is moderate in our experiments.

We also asked the participants whether they would like to receive sociofeedback or not. Out of 20 participants, 19 responded in favor of receiving sociofeedback.

At the end of the experiment, we asked to participants to leave any suggestion that they might have about the experiment. Some participants suggested improvements for the feedback messages. These suggestions were about the timing of the feedback, and also about making the feedback more natural. For instance, one participant suggested the following: “The conversation was interrupted while we were talking happily. When people are having a good conversation, it’s better to use body language only instead of voice”. We intend to work on further improvements of our setup in light of these suggestions.

## 5 Conclusions and Future Work

In this paper, we presented a user study about the Nao robot as social mediator. In the first part of the study, we investigated whether users can understand the feedback messages and compared two modalities for feedback (audio only and audio combined with gestures). In the second part of the study, we assessed how Nao is

**Table 11** Average ratings of each assessment question. Each column shows the ratings for each question, where each row represents a social scenario.

Question	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Normal	4	4	3	2	4	5	4	5
Uninterested	4	4	3	1	4	4	5	4
Overly talkative	4	5	3	3	4	5	5	5
Aggressive	5	5	3	2	4	5	4	5
Total average	4.3	4.5	3	2	4	4.8	4.5	4.8

**Table 12** Average ratings for the Godspeed questionnaire (5-likert scale).

Characteristics	Average Values
<b>Anthropomorphism</b> Machine/human like	4
Moving rigidly/elegantly	3
<b>Animacy</b> Mechanical/organic	4
Inert/interactive	4
<b>Likability</b> Dislike/like	5
Unfriendly/friendly	4
<b>Perceived Intelligence</b> Ignorant/knowledgeable	4
Unintelligent/intelligent	4
<b>Perceived Safety</b> Calm/agitated	3
Quiescent/surprised	3

perceived by people in the role of social mediator in two-person dialogs. In this setting, the sociofeedback system monitored an ongoing conversation, and provided feedback to the participants regarding their social behavior. The feedback was delivered by a humanoid Nao robot which was interfaced with the sociofeedback system. We aimed to investigate how the feedback from the humanoid robot is perceived by humans. To this end, we conducted a survey with 20 participants, where the participants were engaged in a discussion, and the feedback messages were delivered by Nao to the participants. The participants assessed the content, timing, relevance, and their liking of the feedback after receiving each feedback message.

We observed that the participants clearly liked receiving feedback from Nao robot. The agreement scores are very high, showing that the participants agreed with the provided feedback. There is room for improvement in the timing of the feedback. We will try to improve the timing in future experiments.

At the end, each participant assessed the robot in the role of social mediator and rated it on a Godspeed questionnaire. The ratings for all Godspeed criteria are high that implies that participants liked a humanoid robot as a social mediator. Only with regard to perceived safety, the evaluation was only mildly positive; this may be explained by the fact that the robot was sitting near the participants. However, the average rat-

ing is still acceptable, and this issue may not be very critical.

Overall, this study suggests that sociofeedback by the Nao robot can be accurately identified and is appreciated by participants. The findings of this study helped us validate the workings of the sociofeedback system, and it also provided us valuable insight about the human perception of conversation based feedback provided by a robot. We also determined that the best way to provide this feedback is via audio message accompanied by a gesture. In this work we do not use the emotional aspects of the dialogues, but in future work we can determine the emotional aspect of a participant's dialog using his/her speech.

In future work we plan to develop a similar social state estimation module for Nadine robot [41] at Institute for Media Innovation, Nanyang Technological University, Singapore. To this end, we have collected multi-modal (audio and video) dataset and trained the sociofeedback system on the new corpus [42]. Secondly, we will attempt to scale the proposed system to multi-party dialogs. We also intend to further improve the feedback delivered by the robot, and make it look more interactive, so that in future it can be a part of group discussions.

**Acknowledgements** This research is supported by the BeingTogether Centre, a collaboration between Nanyang Technological University (NTU) Singapore and University of North Carolina (UNC) at Chapel Hill. The BeingTogether Centre is supported by the National Research Foundation, Prime Ministers Office, Singapore under its International Research Centres in Singapore Funding Initiative.

## References

1. T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and autonomous systems*, vol. 42, no. 3, pp. 143–166, 2003.
2. H. Li, J.-J. Cabibihan, and Y. K. Tan, "Towards an effective design of social robots," *International Journal of Social Robotics*, vol. 3, no. 4, pp. 333–335, 2011.
3. T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "A communication robot in a shopping mall," *IEEE Transactions on Robotics*, vol. 26, no. 5, pp. 897–913, 2010.
4. S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "Minerva: A second-generation museum tour-guide robot," in *In Proceedings of IEEE International Conference on Robotics and Automation (ICRA 99)*, 1999.
5. F. Tanaka, A. Cicourel, and J. R. Movellan, "Socialization between toddlers and robots at an early childhood education center," *Proceedings of the National Academy of Science*, vol. 104, pp. 17954–17958, Nov. 2007.
6. B. Graf, C. Parltitz, and M. Hägele, "Robotic home assistant care-o-bot<sup>®</sup> 3 product vision and innovation platform," in *HCI (2)*, pp. 312–320, 2009.
7. K. Williams and C. Breazeal, "A reasoning architecture for human-robot joint tasks using physics-, social-, and capability-based logic," in *IROS*, pp. 664–671, 2012.
8. D. François, D. Polani, and K. Dautenhahn, "Towards socially adaptive robots: A novel method for real time recognition of human-robot interaction styles," in *Humanoids*, pp. 353–359, 2008.
9. F. Papadopoulos, K. Dautenhahn, and W. C. Ho, "Exploring the use of robots as social mediators in a remote human-human collaborative communication experiment," *Paladyn*, vol. 3, no. 1, pp. 1–10, 2012.
10. K. Dautenhahn, "Socially intelligent robots: dimensions of human-robot interaction," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1480, pp. 679–704, 2007.
11. D. Bohus and E. Horvitz, "Dialog in the open world: platform and applications," in *Proceedings of the 2009 international conference on Multimodal interfaces*, pp. 31–38, ACM, 2009.
12. A. A. Salah, T. Gevers, N. Sebe, and A. Vinciarelli, *Human Behavior Understanding: First International Workshop, HBU 2010, Istanbul, Turkey, August 22, 2010, Proceedings*, vol. 6219. Springer, 2010.
13. A. A. Salah, J. Ruiz-del Solar, Ç. Meriçli, and P.-Y. Oudeyer, "Human behavior understanding for robotics," in *Human Behavior Understanding*, pp. 1–16, Springer, 2012.
14. U. Rasheed, Y. Tahir, S. Dauwels, and J. Dauwels, "Real-time comprehensive sociometrics for two-person dialogs," in *Human Behavior Understanding*, pp. 196–208, Springer, 2013.
15. A. S. Pentland, *Honest signals*. MIT press, 2010.
16. Y. Tahir, U. Rasheed, S. Dauwels, J. Dauwels, N. Thalmann, and D. Thalmann, "Nao as social mediator: A user study," in *Robots in public spaces: towards multi-party, short-term, dynamic human-robot interaction*, Springer, 2013.
17. C. Bartneck, E. Croft, and D. Kulic, "Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots," in *Metrics for HRI Workshop, Technical Report*, vol. 471, pp. 37–44, Cite-seer, 2008.
18. C. Breazeal and B. Scassellati, "A context-dependent attention system for a social robot," *rm*, vol. 255, p. 3, 1999.
19. C. L. Sidner and M. Dzikovska, "A first experiment in engagement for human-robot interaction in hosting activities," in *Advances in Natural Multimodal Dialogue Systems*, pp. 55–76, Springer, 2005.
20. M. E. Pollack, L. Brown, D. Colbry, C. Orosz, B. Peintner, S. Ramakrishnan, S. Engberg, J. T. Matthews, J. Dunbar-Jacob, C. E. McCarthy, et al., "Pearl: A mobile robotic assistant for the elderly," in *AAAI workshop on automation as eldercare*, vol. 2002, pp. 85–91, 2002.
21. H. Ishiguro, T. Ono, M. Imai, T. Maeda, T. Kanda, and R. Nakatsu, "Robovie: an interactive humanoid robot," *Industrial robot: An international journal*, vol. 28, no. 6, pp. 498–504, 2001.
22. A. Billard, "Robota: Clever toy and educational tool," *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 259–269, 2003.
23. C. D. Kidd, W. Taggart, and S. Turkle, "A sociable robot to encourage social interaction among the elderly," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pp. 3972–3976, IEEE, 2006.
24. K. C. Welch, U. Lahiri, Z. Warren, and N. Sarkar, "An approach to the design of socially acceptable robots for

- children with autism spectrum disorders,” *International Journal of Social Robotics*, vol. 2, no. 4, pp. 391–403, 2010.
25. I. Fujimoto, T. Matsumoto, P. R. S. De Silva, M. Kobayashi, and M. Higashi, “Mimicking and evaluating human motion to improve the imitation skill of children with autism through a robot,” *International Journal of Social Robotics*, vol. 3, no. 4, pp. 349–357, 2011.
  26. G. Schiavone, D. Formica, F. Taffoni, D. Campolo, E. Guglielmelli, and F. Keller, “Multimodal ecological technology: From child’s social behavior assessment to child-robot interaction improvement,” *International Journal of Social Robotics*, vol. 3, no. 1, pp. 69–81, 2011.
  27. J.-J. Cabibihan, H. Javed, M. Ang Jr, and S. M. Aljunied, “Why robots? a survey on the roles and benefits of social robots in the therapy of children with autism,” *International Journal of Social Robotics*, pp. 1–26, 2013.
  28. A. Powers and S. Kiesler, “The advisor robot: tracing people’s mental model from a robot’s physical attributes,” in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pp. 218–225, ACM, 2006.
  29. M. E. Foster, A. Gaschler, M. Giuliani, A. Isard, M. Pateraki, and R. Petrick, “Two people walk into a bar: Dynamic multi-party social interaction with a robot agent,” in *Proceedings of the 14th ACM international conference on Multimodal interaction*, pp. 3–10, ACM, 2012.
  30. “Darpa cognitive agent that learns and organizes (calo) project..”
  31. G. Tur, A. Stolcke, L. Voss, S. Peters, D. Hakkani-Tur, J. Dowding, B. Favre, R. Fernández, M. Frampton, M. Frandsen, *et al.*, “The calo meeting assistant system,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1601–1611, 2010.
  32. S. Al Moubayed, J. Beskow, G. Skantze, and B. Granström, “Furhat: a back-projected human-like robot head for multiparty human-machine interaction,” in *Cognitive Behavioural Systems*, pp. 114–130, Springer, 2012.
  33. J. Harris and E. Sharlin, “Exploring the affect of abstract motion in social human-robot interaction,” in *RO-MAN, 2011 IEEE*, pp. 441–448, IEEE, 2011.
  34. M. Saerbeck and C. Bartneck, “Perception of affect elicited by robot motion,” in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pp. 53–60, IEEE Press, 2010.
  35. A. Beck, A. Hiole, A. Mazel, and L. Cañamero, “Interpretation of emotional body language displayed by robots,” in *Proceedings of the 3rd international workshop on Affective interaction in natural environments*, pp. 37–42, ACM, 2010.
  36. J. Ham, R. Bokhorst, and J. Cabibihan, “The influence of gazing and gestures of a storytelling robot on its persuasive power,” in *International conference on social robotics*, 2011.
  37. A. Delaborde and L. Devillers, “Use of nonverbal speech cues in social interaction between human and robot: emotional and interactional markers,” in *Proceedings of the 3rd international workshop on Affective interaction in natural environments*, pp. 75–80, ACM, 2010.
  38. L. Yu and H. Liu, “Feature selection for high-dimensional data: A fast correlation-based filter solution,” in *Proceedings of the 20th international conference on machine learning (ICML-03)*, pp. 856–863, 2003.
  39. M. A. Hall, “Correlation-based feature selection for machine learning,” 1999.
  40. K. Werner, J. Oberzaucher, and F. Werner, “Evaluation of human robot interaction factors of a socially assistive robot together with older people,” in *Complex, Intelligent and Software Intensive Systems (CISIS), 2012 Sixth International Conference on*, pp. 455–460, IEEE, 2012.
  41. “Nadine robot.,” 2015.
  42. Y. Tahir, D. Chakraborty, T. Maszczyk, S. Dauwels, J. Dauwels, N. Thalmann, and D. Thalmann, “Real-time sociometrics from audio-visual features for two-person dialogs,” in *2015 IEEE International Conference on Digital Signal Processing (DSP)*, pp. 823–827, IEEE, 2015.