Int. J. Human-Computer Studies 80 (2015) 14–23

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

# "C'Mon dude!": Users adapt their behaviour to a robotic agent with an attention model ☆

Lawrence Cavedon [a,*], Christian Kroos [b,d], Damith Herath [b,e], Denis Burnham [b], Laura Bishop [b], Yvonne Leung [b], Catherine J. Stevens [b,c]

[a] School of Computer Science and IT, RMIT University, GPO Box 2476, Melbourne 3001, Victoria, Australia
[b] MARCS Institute, University of Western Sydney, Australia
[c] School of Social Sciences and Psychology, University of Western Sydney, Australia
[d] Alternate Anatomies Lab, Curtin University, Perth, Australia
[e] Robological Pty Ltd, Chatswood, NSW, Australia

ARTICLE INFO

*Article history:*
Received 16 May 2014
Received in revised form
2 December 2014
Accepted 23 February 2015
Communicated by C. Bartneck
Available online 3 March 2015

*Keywords:*
Human–robot interaction
Attention model
Social interaction
Evaluation
Engagement

ABSTRACT

Social cues facilitate engagement between interaction participants, whether they be two (or more) humans or a human and an artificial agent such as a robot. Previous work specific to human–agent/robot interaction has demonstrated the efficacy of implemented social behaviours, such as eye-gaze or facial gestures, for demonstrating the illusion of engagement and positively impacting interaction with a human. We describe the implementation of THAMBS, The Thinking Head Attention Model and Behavioural System, which is used to model attention controlling how a virtual agent reacts to external audio and visual stimuli within the context of an interaction with a human user. We evaluate the efficacy of THAMBS for a virtual agent mounted on a robotic platform in a controlled experimental setting, and collect both task- and behavioural-performance variables, along with self-reported ratings of engagement. Our results show that human subjects noticeably engaged more often, and in more interesting ways, with the robotic agent when THAMBS was activated, indicating that even a rudimentary display of attention by the robot elicits significantly increased attention by the human. Back-channelling had less of an effect on user behaviour. THAMBS and back-channelling did not interact and neither had an effect on self-report ratings. Our results concerning THAMBS hold implications for the design of successful human–robot interactive behaviours.



## 1. Introduction

Interaction between two interlocutors, whether via spoken language or otherwise, is heavily impacted by social cues, whether they be verbal or non-verbal. Communicative mechanisms for such social behaviours include back-channels such as head nods and "yeah" verbal utterances (Hess and Johnston, 1988; Hiroji, 2006), eye-gaze (Langton et al., 2000), body position and gestures (Argyle and Cook, 1976; Goldin-Meadow, 1999; Langton and Bruce, 2000; McNeill, 2005).

The influence of social cues on (human) interaction has been shown to extend to settings where one of the participants is an artificial agent, such as an animated *embodied conversational agent* (*ECA*) or a robot. For example, Colburn et al. (2000) demonstrated that an eye gaze model in an avatar induces changes in eye gaze of a human conversing with the avatar. The eye gaze model helps direct attention to the speaker and indicates to participants who is speaking and who is listening. Eye gaze in particular can direct a user's gaze in a complex virtual scene (Bailly et al., 2010). Embodied conversational agent REA uses eye gaze together with body posture, hand gestures, and facial expression to regulate conversation (Cassell et al., 1999). In a recent study, a Honda humanoid robot was evaluated more positively when hand and arm gestures were displayed along with speech and participants were asked to direct their attention to the robot during the interaction (Salem et al., 2011). Back-channelling behaviours can influence communicative-task success (e.g., Bavelas et al., 2000); however, real-time back-channel feedback that is related to the content of dialogue is computationally challenging (Jonsdottir et al., 2007).

In this paper, we describe an experiment involving human interaction with a robot that incorporates a model of "attention": i.e., we have provided our robot ("Segie") with a parameterised model of attention to the (human) interaction participants. THAMBS, the *Thinking Head Attention Model and Behavioural System*, influences the level to which Segie displays "engagement"

☆This paper has been recommended for acceptance by C. Bartneck.

* Corresponding author. Tel.: +613 9925 2325.
*E-mail address:* lawrence.cavedon@rmit.edu.au (L. Cavedon).

http://dx.doi.org/10.1016/j.ijhcs.2015.02.012
1071-5819/

within an interaction, or conversely, how easily Segie is distracted from an interaction by external influences. THAMBS is embedded in the *Thinking Head architecture* (Herath et al., 2010) that controls Segie, and is connected to Segie's perceptors and actuators, including: auditory perception (speech recognition and long-range directional) and speech synthesis; visual perception via mounted cameras; mobility; and control of a monitor displaying an animated human head.[1]

Our experiment explores the influence of the attention model, i.e., THAMBS, on the behaviour of the human participants. In particular, we observe that the more predictable or overt display of attention by the robot within the interaction elicits direct and explicit attempts to engage Segie—this occurs to a significantly greater degree when THAMBS is in operation as compared to when THAMBS is off or when the robot's attention behaviour is more random. Moreover, the influence of standard social behaviours, such as eye-gaze, is more pronounced (to a significant degree) when THAMBS is on.

In the following section, we summarise previous findings regarding influence of social communicative behaviours on interaction participants, for both human–human and human–agent/robot communication. We then describe our robot platform, the Thinking Head architecture that drives it, and the THAMBS model of attention and behavior. This is followed by a description of our evaluation approach, involving a controlled experimental design, with task and behavioural (video-coded) performance metrics, as well as self-reported ratings of engagement by subjects. We then present the results and discuss how they support our claims related to human–robot interactive behavior. We close by summarizing our conclusions and pointing to proposed future extensions of this work.

## 2. Social interactive behaviours in human–human/robot/agent interaction

*Human–Robot Interaction* (*HRI*) is a relatively young field of research investigating the nature of the use of robotic systems “by or with humans" (Goodrich and Schultz, 2007). A key emerging topic within that field is the nature of socially interactive robots (Fong et al., 2003): i.e., robots which exhibit some general characteristics to facilitate social aspects of interaction, either with other robots or with humans in their environment. Fong et al. identify a number of “human social” characteristics for HRI, including: expressing and/or perceiving emotion; using natural cues (gaze, gestures, etc.); and establishing/maintaining social relationships (Fong et al., 2003). Much of the focus in HRI research has been on designing robots capable of displaying social behaviours, particularly by visualising emotional reactions (e.g., through facial expressions; Breazeal, 2002), using gaze to specify focus of attention (Imai et al., 2003; Sun et al., 2008), or via other aspects of physical embodiment to collaborate with a human participant. For brevity, we focus here on social aspects of verbal communication between two participants, and in particular, those aspects that indicate *attention to the other participant and their communicative actions*.

There are many verbal and non-verbal strategies and behaviours used to manage interaction in human–human communication. Some of these play important functional roles in two- or multi-party interaction, for example, managing turn-taking (Starkey, 1972). We are specifically interested here in behaviours that increase and/or indicate “engagement”, and in particular are non-verbal or otherwise not related to the information content of the interaction.

One of the most studied of such phenomena is the role of eye-gaze, in particular, in a listener holding gaze with a speaker (although the reverse is also important). The importance of the role of eye-gaze in dialogue and interactive communication has been extensively reported for simulating and maintaining attention and engagement, as well as for functional phenomena such as indicating turn (or maintaining it). There have been a number of implementations of gaze models in robots and ECAs, as well as evaluations of their effect on human conversational participants (e.g., Breazeal and Scassellati, 1999; Colburn et al., 2000; Mutlu et al., 2006; Peters and Itti, 2006; Bailly et al., 2010; Bee et al., 2010; Zhang et al., 2010).

Another important social phenomenon, once again from the listener, is *back-channel feedback*: this may be verbal (for English, examples include “*yeah*”, “*uh huh*”, and “*mhm*”) or non-verbal (e.g., head nods or facial expressions). Back-channel feedback is used by listeners (often unconsciously) to indicate to a speaker that they are listening and paying attention, often in response to specific prosodic cues from the speaker (Ward and Tsukahara, 2000). Such feedback can play a critical role in the success of dialogue tasks. For example, (Bavelas et al., 2000) found that a lack of back-channel feedback from the listener impacted story-tellers, who would falter and fail to finish telling their story. Rajan et al. (2001) similarly observed a high incidence of back-channelling in successful tutor–student dialogues.

We investigate back-channelling behaviour in the experiment below. However, while a number of sophisticated/complex models of back-channelling have been proposed and implemented in animated conversational agents (Fujie et al., 2005; Kopp et al., 2008; Kopp et al., 2007; Poggi, 2005), we use an approach whereby the robot's back-channelling behavior is controlled by the (human) Wizard—our main interest is in how the phenomenon of back-channelling is independent of or interacts with the THAMBS attention model described in the following section.

While such behaviours that *indicate* or inspire attention/engagement have been extensively studied in the literature on Conversational Agents and HRI, there has been relatively little work involving *cognitive* models of attention *to an interaction and its participants*, i.e., models of processes whereby an agent “decides” where to direct attention within the context of an interaction, with the notable exception of the work of Breazeal and Scassellati (1999). Greater focus within HRI has been placed on developing models of *Joint Attention*, i.e., mechanisms by which dialogue participants draw focus to an environmental object typically being referenced by the dialogue (e.g., Imai et al., 2003). The THAMBS Model of Attention described in the following section addresses different aims to such work, by firstly considering all external stimuli—both visual and auditory, those related to the dialogue as well as those extraneous to it—and developing a cognitive-level model of how the robotic agent is influenced by these in the context of a conversation between the robot and a human participant. We claim that such a model leads to more natural *social* conversational behavior, which in turn facilitates more successful communication patterns between the robot and the human.

## 3. The robotic platform and the THAMBS attention model

Segie, the robot participant in the interaction experiments described below, consists of a Segway Robotic Mobility Platform[2] with a range of perceptual capabilities, actuators that drive its mobile and

[1] Other physical incarnations of the Thinking Head include an installation of the monitor display on a robotic arm that swivels to visually focus the animated head on a source of attention or distraction (Herath et al., 2010).

[2] http://rmp.segway.com/

interactive behaviours, and a purpose-built software platform—the Thinking Head platform—for integrating capabilities and controlling all behaviours, including that of an animated head mounted on the robot platform. The attention model—THAMBS (Thinking Head Attention Model and Behavioural System; Kroos et al., 2012)—is the most salient aspect of the Thinking Head software infrastructure for the focus of this paper and is described in most detail below. Further details of aspects of the software and robotic platforms can be found in Herath et al. (2010).

### 3.1. The thinking head platform

The Thinking Head (TH) platform is a flexible software architecture for a conversational agent rendered as an animation of the head of the performance artist Stelarc (see Fig. 1). The motivation driving the implementation of the TH platform was to develop a flexible "plug-and-play" software infrastructure for a conversational agent, allowing components implementing specific capabilities (e.g., speech recognition, dialogue management) to be easily replaced with similarly-capable components to allow experimentation and evaluation within a real-time interactive setting. All capabilities (perceptors, processors, and actuators) communicate via a flexible event-driven communication framework designed to facilitate real-time interaction between components—see Herath et al., (2010) for details.

The Thinking Head framework, and in particular the robotic instantiation used in our experiments, provides for a number of standard sensor, actuator, and audio-visual devices to facilitate human–robot interaction. In the simple configuration used here, speech recognition and speech-synthesis—using off-the-shelf components (e.g., Dragon Naturally Speaking)—facilitate basic spoken interaction; dialogue management is implemented by a chatbot architecture enhanced with task-specific rules; and the (purpose-built) animated 3D head has its lip-movements synchronised to the synthesised speech and is capable of displaying basic emotional visual expressions.

Other capabilities were implemented specifically for the purpose of experimenting with models of attention on a robotic platform:

- *Auditory localisation* provides accurate information on the instantaneous locations (azimuth) of multiple moving interlocutors in a noisy and reverberant environment.
- *Visual tracking*: This includes ***people and object tracking***, which provides localisation and height of all people and standing objects within the mounted camera's field of vision; and ***face tracking***, which is capable of detecting and then continuously tracking a single face within the camera's field of vision.

These capabilities are particularly salient to our experiments with the attention model: while the visual tracking provides the semblance of physical "engagement" with a human subject, auditory and visual localisation drives a "distraction" behaviour, where a noise or another object diverts the robot's attention away from the interaction.

### 3.2. Robotic platforms

The original robotic embodiment for the Thinking Head was a computer monitor (displaying the 3D animated head) mounted on a robotic arm with six degrees of freedom of movement (see Kroos et al., 2011). In the experiments described here, we used a simpler robotic platform, designed to constrain the variability of behaviours while retaining sufficient mobility to display clear differences in directed attention. The Thinking Head infrastructure was mounted on a Segway Robotic Mobility Platform, a simple robust mobile robot platform with an open programming interface. An LCD screen mounted on the robot platform displays the avatar face and a Kinect device is used for audio and visual perceptual input: see Fig. 2. This implementation lacks the compelling presence of the implementation on the robotic arm, but provides the basic mobility and perceptual capabilities required for our experiments, without the potentially confounding factor of the more complex and compelling implementation. We argue that the observed reactions from the human subjects can be more reliably ascribed to the social behaviours than to any aspect of the physical characteristics under this simpler implementation.

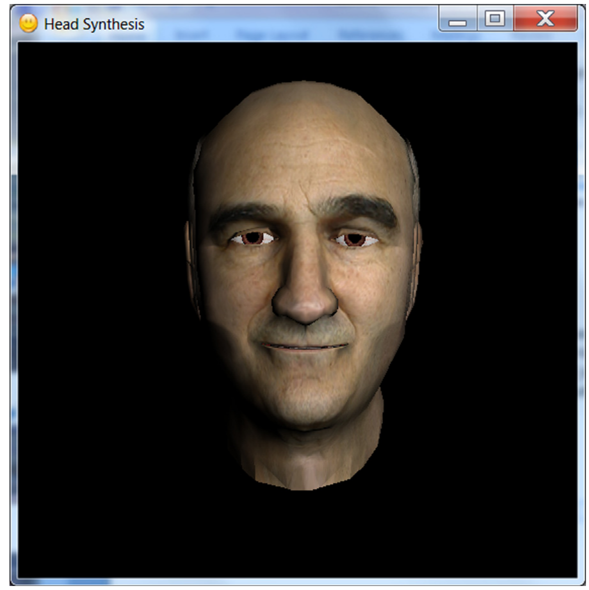


**Fig. 1.** Animated head.

### 3.3. THAMBS: the thinking head attention model and behavioural system

The human–robot interfaces described in the previous section provide the robot platform with information about its environment. We argue that consistent interactions with humans emerge only if the robot's sensing capabilities of the environment are related to its motor capabilities in a meaningful way and according to the expectation of the human user. Only a tight coupling between perception and action can generate behaviour that convincingly creates the illusion that the robot is an intentional agent with its own agenda and with this enables a different quality of human–robot interaction. The Thinking Head Attention Model and Behavioural System (THAMBS) was developed against this background.

THAMBS is a perception-action control architecture that consists of the following high-level modules: (1) a perceptual system; (2) an attention system; (3) a central control system; and (4) a motor system.

The perceptual system wraps the lower level sensing streams and creates within-system standardised perceptual events. For the current experiment only two sensing abilities of THAMBS were activated: acoustic source localisation and visual people detection in 3D space. The perceptual system has its own set of thresholds acting–for instance, on confidence values returned by the sensing systems–and also computes the deltas for each input ("velocities"). These thresholds were determined empirically to filter background stimuli from being considered to be significant attention-attracting events. This was done (once only) by having two experimenters judge whether the robotic agent was reacting to irrelevant stimuli (thresholds set too low) or ignoring pertinent stimuli (thresholds set too high). For example, the thresholding

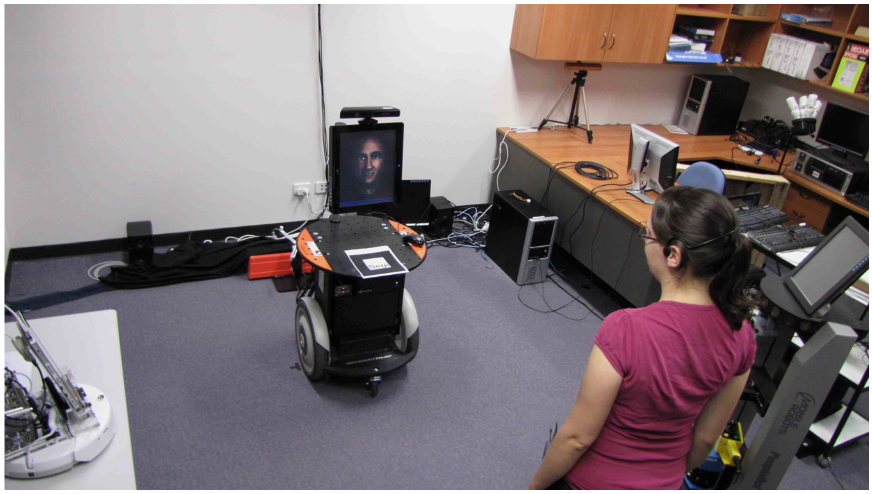

**Fig. 2.** Segie—the software-rendered head is displayed on an LCD screen affixed to a Segway unit.

mechanism meant that a static object was not considered a new attention focus based on velocity despite drift in the tracking or slight body sway.

The generated perceptual events are passed on to the attention system. Algorithmic attention models have been studied for some time. The majority of these are biologically inspired (Peters and Itti, 2006; Bosse et al., 2006; Sun et al., 2008) and have been only applied in computer agents acting in a virtual environment (Kim et al., 2005) bypassing the difficult task of real world object recognition (but see Breazeal and Scassellati, 1999). The identity of objects placed in a virtual environment can be made known to the attention model of the agent; this option is typically not available when dealing with a robot and real world sensing.

The attention system of THAMBS checks the generated perceptual events individually against attention thresholds specific to each type of perceptual event. Those events that have values below the threshold are considered "subliminal" and can be still further processed but are never passed to the central control system. Note that there are thresholds for many aspects of the perceptual event, e.g., a high velocity of an otherwise sub-threshold event can allow it to become attended. Perceptual events that pass the test create an attention focus that currently is entirely spatially organised ("pay attention to region X"), thus, identity of two foci is assumed if they refer to the same spatial region.

The attention system then assigns an initial weight and an exponential decay function to the focus based on the current task priorities specified by the central control system. These depend, of course, on the overall state of the robotic Thinking Head with respect to the ongoing interaction and its ultimate goal. The attention system determines a single attended event from all available foci using a winner-takes-all strategy and relays it to the central control system as the presently attended event. It also directly generates a motor goal to bring the attended event at the centre of the robot's mobile visual system and forwards this motor goal directly to the motor system.

The central control system evaluates the attended event based on the values of a larger set of THAMBS state variables and generates a behavioural response. In the experiments described here, this consisted of a single simple action, but in general may include a pre-defined temporal sequence of motor events to implement a potentially complex action: e.g., "direct monitor to point (*x*,*y*,*z*)"; raise eyebrows; say "Hello". The behaviour trigger is currently realised in the form of conditional rules acting on various thresholds. If the attended event is sufficiently close (varying thresholds) to one of the pre-defined behavioural responses, then this behaviour i.e., a temporal sequence of motor goals, is considered an appropriate response and is activated.

The abstract motor goals (e.g., "follow person with id 2") are transformed into sequences of implementation-specific motor primitives by the motor system. The set of motor primitives covers both movements of the robot and facial movements of the ECA displayed on the monitor. Motor goals coming from the central control systems will suppress goals from the attention system, unless the latter have an associated weight higher than task-specific threshold. Technical details can be found in Kroos et al. (2011).

## 4. Rationale and research questions

Previous investigations of avatar or robot nonverbal, gestural feedback indicate that cues or responses such as head turns are effective when they occur at meaningful rather than random times or points in the interaction (e.g., Yamazaki et al., 2008). As in human to human interaction, back-channelling provides feedback on a number of levels simultaneously (Brunner, 1979), such as providing information and cues to emotion and attitudes (Kopp et al., 2008). Back-channelling enables one member of the interacting dyad or group to know the partner's state and to feel comfortable to speak (Fujie et al., 2005). Where the robot or avatar provides feedback and cues, humans will adapt their behaviour (Zhang et al., 2010).

In the present experiment, we investigate the effect of an attention model and back-channelling on the quantity and quality of user interaction. Back-channelling was in the form of verbal utterances that were produced by robot Segie, driven by a Wizard of Oz set-up. The $3 \times 3$ between-subjects factorial design comprised the independent variables attention model or THAMBS with three levels (*on, random, off*) and back-channelling by Segie with three levels (*on, random, off*). The inclusion of a *random* setting is to ascertain whether user behaviour is not influenced simply by the presence of attention and/or back-channelling but maximally influenced when systems are on and working meaningfully than when on but working in a random (non-meaningful) way. Effectiveness of the attention model and back-channelling will be inferred from the user's behaviour including the number of comparisons they make between two robots in the room (the goal of the task participants are given), the number and kinds of interactions they have with Segie, and explicit ratings assigned to 10 Likert scales regarding the user's judgments of engagement with Segie, their liking of Segie, and so on. Specifically, the dependent variables were the number of *pauses* that participants made as they waited for Segie to orient; the number of times participants attempted to *re-engage Segie's attention*; the number of times eye contact was made between Segie and participants; and the number of comparisons made between the two robots.

The experimental design links our questions to observable phenomena as follows: (1) if THAMBS works effectively, i.e., its parameters are appropriate, then users will engage more when THAMBS is on than when THAMBS is off or random; (2) if human-to-human style back-channelling appeals to users then they will engage more with the robot when back-channelling is on than when it is off or random; and (3) investigates the independence or interactions between back-channelling and THAMBS.

## 5. Method

### 5.1. Participants

One hundred and twenty undergraduate psychology students at the University of Western Sydney (age $M=21.9$, $SD=5.9$; 98 female) completed the experiment. All participants reported normal or corrected to normal vision and two reported corrected hearing. Thirteen reported a first language other than English (Afghan, Arabic, Assyrian, Hindi, Hmong, Mandarin, Russian, Thai, Zulu). Two had participated in a prior experiment involving interaction with robots, but none reported any other experience with robots. Participants received course credit and a $10 reimbursement (see Section 5.5).

All data from seven participants were excluded because they did not follow the task instructions, describing the two robots individually instead of comparing them. All data from four other participants were excluded due to technical problems resulting in incomplete video recording. The sample used in the data analysis therefore totalled 109 with 12 participants in each of 8 conditions and 13 participants in the THAMBS-*on*, Back-channelling-*off* condition.

### 5.2. Robot platforms

Participants interacted with Segie, the Segway platform with the LCD screen displaying the 3D avatar mounted on it, described in Section 3.2. This is one of four different robotic embodiments of the Thinking Head using the Thinking Head Attention and Behavioural System (THAMBS) (Kroos et al., 2012). The experimental task involved comparing and contrasting the physical features of two other robots: “Peoplebot” and “Roomba” (see Fig. 3). Peoplebot comprises a touchscreen mounted on a four-foot tall grey frame with wheels. To provide participants with additional features to compare during the task, a yellow joystick and power pack sat on the base of the robot during the experimental trials, and a toy flower was affixed midway up its frame. The Roomba comprises an Asus tablet mounted on a white iRobot Create robot platform (the research version of the Roomba vacuum robot). Peoplebot stood on the floor during the experimental sessions and Roomba sat on a table. Both remained switched off.

### 5.3. Equipment

Segie was linked to a PC, and its speech and movement settings were controlled by the experimenter via the THAMBS graphical user interface in MATLAB (The MathWorks, Inc.). Two video recordings were made of each experimental session, one from the robot's perspective (i.e. front-view) and one from the back of the room (i.e. rear-view). Video data were collected on two Grass Valley Turbo iDDR recording units. Participants were asked to wear a Sennheiser EW 300 G2 wireless microphone headset (see Fig. 4), and audio recordings of their speech and the speech produced by the robot were made on one of the iDDR recording units.

Experimental sessions were conducted while the participant was alone in the robotics lab. The experimenter initiated audio and video recordings and manipulated Segie's speech and movement settings from an adjacent control room. The front-view video captured by the robot was visible to the experimenter at all times on a computer monitor. The experimenter used an Asus tablet running IBM Viavoice text-to-speech synthesis software to control the robot's speech, while THAMBS autonomously controlled all other aspects of the robot's behaviour (i.e., movement, focus of attention).

At the end of the experimental session, participants completed demographic and feedback questionnaires using an Asus tablet with a keyboard and mouse attached.

### 5.4. Task and stimuli

An interaction task was designed that involved individual participants moving around a room to look at and describe to Segie features of two robots in the room. A task that involved movement around the room was necessary so that the tracking capacity of Segie when THAMBS was active would be apparent. One robot was an Adept Mobilerobots Peoplebot and the other was an iRobot Roomba robot that had been modified to include a computer tablet as a display. Participants were asked to introduce the two robots to Segie by verbally comparing and contrasting the Peoplebot and Roomba physical features. Participants were asked to make as many comparisons as possible during this period, with a suggested (challenging) goal of 15. They were told that Segie would be tested at the end of three minutes to see if he had learned the difference between the Peoplebot and Roomba robots. To encourage participants to focus on communicating with Segie and to make as many comparisons as possible, they were told that they would earn $10 if Segie successfully identified the two robots they had been describing.

The task was designed to ensure that it was challenging and engaging and difficult enough to avoid ceiling effects. A challenging task of making comparisons was chosen for two main reasons. First, the comparison aspect ensured that participants moved between the two robots that they were describing to Segie so as to engage the THAMBS tracking of participant movement. Second, robots were used as they were unfamiliar objects to participants which made the task challenging, but it also meant that many of the features on the robots were not easily labelled, i.e., participants would not know what to call some of the features. Thus, participants described features in their own words, and the number of comparisons, rather than features, was tallied.

In the back-channelling-*on* condition, a Wizard of Oz design was used: the experimenter initiated each of Segie's verbal responses (e.g. “yes”, “okay”, “I see”) by clicking on items in a pre-programmed list of phrases. Segie was made to respond after each comparison to make it seem as though he was acknowledging and comprehending the participant's speech. In the back-channelling-*off* condition, Segie said nothing during the three-minute testing period. In the back-channelling-*random* condition, Segie made verbal responses from a list of phrases according to a pre-programmed timing algorithm, and appeared to vocalise at random, with no regard to what the participant was saying. The interval durations between back channelling responses were generated by drawing random values from a normal distribution centred on 12 s with a standard deviation of 5 s. The mean of 12 s was determined by dividing the duration of the testing period by 15, the number of the requested comparisons. The generated value (the time until the next utterance) was cut to a lower and upper

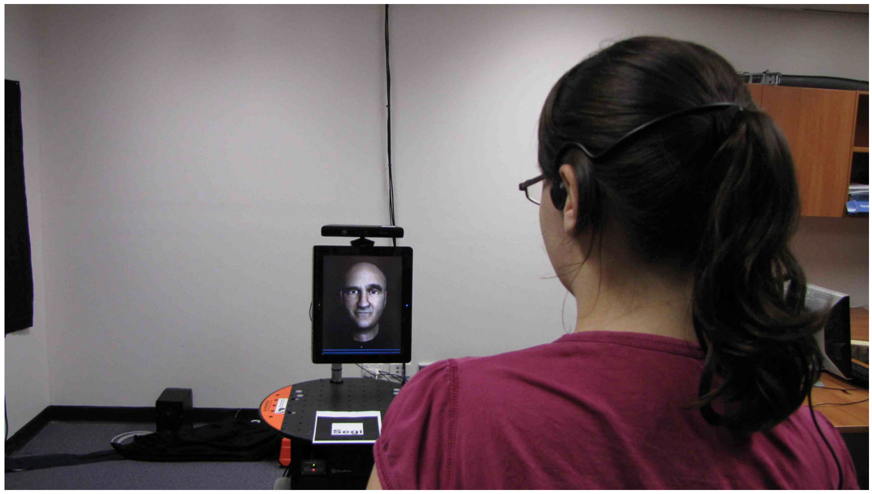

**Fig. 3.** Participants wore a Sennheiser wireless microphone headset for the recording of participant utterances. Speech produced by the robot was recorded into a separate time-locked iDDR recording unit. Video of the interaction was recorded from both front-view and rear-view cameras.

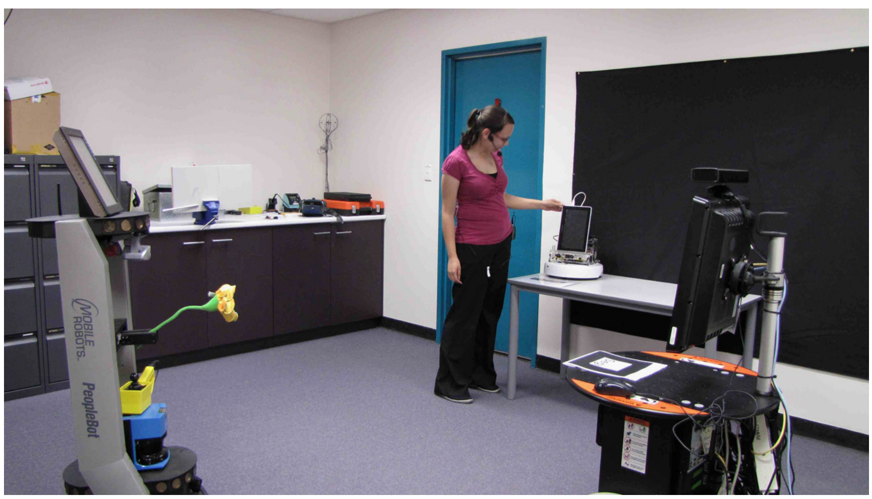


**Fig. 4.** The participant's task was to communicate to Segie similarities and differences between the two other robots (Peoplebot and Roomba) in the room. The robots were on opposite sides of the room so that participants had to walk to either robot to study one and make a comparison with the other robot. In this photo, THAMBS is off.

bound if it fell short of the former or exceeded the latter (set to 2 s and 22 s and respectively). The standard deviation and limit values were determined from the Wizard's interactions with participants in pilot runs testing the technical set-up. Two lists of verbal responses were constructed for use both in the back-channelling-*on* and back-channelling-*random* conditions. The lists included the same set of phrases, but in the reverse order, and were randomly assigned to an experiment run.

During the THAMBS-*on* condition, Segie tracked participants' movement and periodically re-oriented himself to face them by pivoting in place. In the THAMBS-*random* condition, he periodically re-oriented himself according to a pre-programmed algorithm, again by pivoting in place with no regard to where the participant was at the time. The movement values consisted of azimuth target values drawn from a normal distribution centred on 0 (orienting straightforward toward participant in initial position) with a standard deviation of $20^{\circ}$ and an lower and upper cut-off limit of $-40^{\circ}$ and $40^{\circ}$. In the THAMBS-*off* condition, Segie remained entirely still.

### 5.5. Procedure

Participants provided written informed consent and were given verbal instructions at the start of the experimental session. They were then introduced to Segie and told that their task would be to introduce two other robots (Peoplebot and Roomba) to Segie by verbally comparing and contrasting their physical features. The three robots were arranged in a triangular formation in the robotics lab at the University of Western Sydney. Peoplebot and Roomba were oriented with their screens angled slightly towards Segie. Participants began the session standing midway between Peoplebot and Roomba, facing Segie. They were asked to walk between Peoplebot and Roomba as they made their comparisons and to point out each feature individually. They were also instructed to maintain eye contact with Segie and to try to communicate with him as though conversing with another person. Participants were tested individually. They were given a demonstration and then completed a practice trial with verbal feedback from the experimenter. The experimenter then left the room, leaving the participant alone with the three robots. In all conditions, Segie verbally instructed participants to begin at the start of the three-minute testing period and to stop at the end.

The interaction period was three minutes. Participants were asked to make as many comparisons as possible during this period, with a suggested goal of fifteen. They were told that Segie would be tested at the end of the three-minute period to see if he had learned the difference between Peoplebot and Roomba. To encourage participants to focus on communicating with Segie and to make as many comparisons as possible, they were told that they would earn $10 on top of their course credit if Segie successfully identified the two robots they had been describing.

Following the testing period, participants showed Segie a page with photos of five different robots on it, including Peoplebot and Roomba. Participants read out a pre-scripted sentence that instructed Segie to guess which two robots they had been describing. Segie (via the Wizard) always guessed correctly, and participants were rewarded with their bonus $10. Participants then completed a brief demographic questionnaire and gave feedback about their thoughts and experience with Segie. Using a five-point rating scale, they were asked to evaluate the robot's likability, whether Segie kept their attention, whether they would like to interact with Segie again, whether they enjoyed the interaction, whether they thought Segie attended to their speech, and whether they found Segie engaging and humorous. An additional two questions were included for participants in the back-channelling-*on* and back-channelling-*random* conditions: participants were asked to evaluate how comprehensible Segie's speech was and rate whether they felt Segie was speaking just to them.

Once they had completed the feedback questionnaire, participants were debriefed about the aims of the experiment.

## 6. Results

A number of dependent measures were computed. The first three were derived from the video recordings of participant behaviour during the experimental session and refer to engaging Segie; videos were scored by two independent observers. The fourth refers to participant performance on the comparison task.

1. *Pauses*: A point was given if the participant paused at any time and waited for Segie to finish orienting towards them or stopped and waited when he failed to orient towards them. No points were given if they never did this during the trial.
2. *Re-engaging attention*: A point was given each time the participant attempted to attract Segie's attention when he failed to orient towards them. What the person said or did was noted.
3. *Eye contact*: This was scored using the front-view video during the THAMBS conditions by counting the number of times two-way eye contact was made (i.e. participant and robot were looking at each other) after Segie re-oriented. If the participant glanced up more than once while in the same spot and Segie did not move, that only counted as one instance of two-way eye contact.
4. *Number of comparisons*: One point was given for each comparison made between the two robots. Points were given only for *comparisons between robots*, not for descriptions of individual robots—e.g. if a participant pointed out a characteristic of Peoplebot's base, then commented on Roomba's base, they received a point; if they pointed out a characteristic of Peoplebot's base, then commented on Roomba's screen, they did not receive a point. If they repeated something they said earlier, they did not get a second point for the repeated comparison.

Descriptive statistics for the behavioural measures are shown in Table 1. Pauses, attention and eye contact were not computed in the THAMBS-*off* condition as there was no stimulus cue to which such measures could be related; this is reflected in the degrees of freedom in the statistics reported below. Table 2 shows descriptive statistics for the proportion of time participants tried to attract Segie's attention of the total time that Segie was not oriented towards them ("Attract Proportion"), and the proportion of time the participant looked at Segie out of the total number of times Segie reoriented to look at the participant ("Eye Contact Proportion"), as a function of THAMBS and Back-Channelling. Explicit, self-report ratings of engagement, liking, understanding, lifelikeness, humour, etc. were also obtained.

**Table 1**
Means and standard deviations (in brackets) for the four behavioural dependent variables as a function of THAMBS and back-channelling.

| | Pauses | Re-engaging attention | Eye contact | No. of comparisons |
|---|---|---|---|---|
| THAMBS | | | | |
| On | .70 (.74) | .40 (.42) | .90 (.17) | 8.8 (2.8) |
| Random | .42 (.55) | .11 (.13) | .76 (.21) | 9.1 (2.0) |
| Off | n/a | n/a | n/a | 10.1 (2.3) |
| Back-channelling | | | | |
| On | .33 (.70) | .21 (.28) | .84 (.20) | 9.9 (2.2) |
| Random | .79 (.66) | .29 (.41) | .82 (.25) | 8.8 (2.6) |
| Off | .56 (.58) | .26 (.34) | .84 (.16) | 9.3 (2.4) |

**Table 2**
Means and standard deviations (in brackets) for the proportion of time participants tried to attract Segie's attention of the total time that Segie was not oriented towards them ("attract proportion"), and the proportion of time the participant looked at Segie out of the total number of times Segie reoriented to look at the participant ("Eye Contact Proportion"), as a function of THAMBS and back-channelling.

| | Attract proportion | Eye contact proportion |
|---|---|---|
| THAMBS | | |
| On | .36 (.43) | .89 (.18) |
| Random | .11 (.14) | .77 (.21) |
| Off | n/a | n/a |
| Back-channelling | | |
| On | .22 (.32) | .83 (.21) |
| Random | .29 (.41) | .83 (.23) |
| Off | .19 (.28) | .83 (.16) |

All data—video-based scores, number of comparisons, and ratings—were analysed separately using between subjects analyses of variance (THAMBS $\times$ Back-channelling). Correlations between ratings and video-based scores of user behaviour, and correlations between video-based scores were also calculated.

### 6.1. Effects of THAMBS on user behaviour and ratings

The first research question addressed the effectiveness of THAMBS on user behavior. There was a main effect of THAMBS on participants re-engaging Segie's attention, $F(1,72)=15.81$, $p<.001$, Cohen's $d=.93$, with significantly more instances of people attracting Segie's attention when THAMBS was on (Mean ($M$)$=.40$, $SD=.42$) than when THAMBS was random ($M=.11$, $SD=.13$). Participants re-engaged Segie's attention in the following ways: Waving their arms; moving back into his field of vision; clicking their fingers; repeating the previous sentence; saying: "Oh, where are you going, Segie?" or "Over here, Segie!" or "Come on, dude."

A significant main effect of THAMBS was found in participant eye contact with Segie, $F(1,72)=9.65$, $p=.003$, Cohen's $d=.73$. Mean eye contact when THAMBS was on was .90 ($SD=.17$) compared with .76 ($SD=.21$) when THAMBS was random.

There was no significant main effect of THAMBS on the number of pauses that participants made, $F(1,72)=3.7$, $p=.058$. The mean number of pauses when THAMBS was on was .70 ($SD=.74$) compared with .42 ($SD=.55$) in the random condition.

There were significantly fewer comparisons made when THAMBS was on ($M=8.8$, $SD=2.8$) than when THAMBS was off ($M=10.1$, $SD=2.3$), $F(1,100)=5.14$, $p=.026$, Cohen's $d=.51$. It should be noted that there was no main effect of THAMBS on

**Table 3**
Mean ratings of liking, enjoyment and engagement in interacting with Segie under THAMBS and back-channelling (B/C) conditions; minimum possible rating is 1 (“totally disagree”) and maximum possible rating is 5 (“totally agree”).

| | THAMBS-on | | | THAMBS-random | | | THAMBS-off | | |
|---|---|---|---|---|---|---|---|---|---|
| Item | B/C-on | B/C-rand | B/C-off | B/C-on | B/C-rand | B/C-off | B/C-on | B/C-rand | B/C-off |
| I find Segie likeable | 4.3 | 3.9 | 3.9 | 4.3 | 3.8 | 4 | 4.3 | 4.3 | 4.0 |
| I find Segie engaging | 3.7 | 3.9 | 3.4 | 4.1 | 3.8 | 3.9 | 4.2 | 4.1 | 4 |
| I find Segie easy to understand | 4.2 | 4.1 | n/a | 4.5 | 3.9 | n/a | 4.4 | 4.4 | n/a |
| I find Segie life-like | 3.6 | 3.7 | 3.3 | 3.7 | 3.6 | 3.7 | 4.4 | 3.9 | 3.5 |
| I find Segie humorous | 3.2 | 3.7 | 3.2 | 3.6 | 3.8 | 3.5 | 3.6 | 3.3 | 3.3 |
| Segie kept my attention | 3.7 | 4.2 | 4.0 | 3.9 | 4.0 | 4.0 | 3.9 | 3.5 | 4.0 |
| I would like to interact with Segie again | 4.0 | 3.8 | 3.1 | 4.2 | 4.0 | 4.1 | 4.2 | 3.8 | 3.8 |
| I enjoyed interacting with Segie | 4.2 | 4.1 | 3.9 | 4.3 | 4.3 | 4.3 | 4.6 | 4.2 | 4.0 |
| I felt as if Segie was speaking just to me | 4.3 | 3.6 | n/a | 4.1 | 4.1 | n/a | 4.4 | 3.9 | n/a |
| Segie paid attention to what I said | 3.9 | 4.1 | 3.9 | 4.3 | 4.3 | 3.8 | 4.3 | 4.3 | 4.2 |

**Table 4**
Correlations for video-scored dependent variables.

| | Number of comparisons | Pauses | Attracting attention | Eye contact | Interruption |
|---|---|---|---|---|---|
| Number of comparisons | | | | | |
| Pauses | −.27 | | | | |
| Attracting attention | −.44** | .39** | | | |
| Eye contact | −.14 | .26 | .39** | | |
| Interruption | −.30 | −.02 | .07 | .08 | |
| Congratulations | −.05 | .11 | .05 | −.22 | .04 |

** $p < .001$.

the number of comparisons made, $p = .14$. THAMBS had no effect on the number of comparisons made when THAMBS was *on* compared with when it was *random*, or when THAMBS was *random* compared with when it was *off*.

Mean ratings to the 10 rating scale items are shown in Table 3. Two-way (THAMBS × Back-channelling) between-subjects analyses of variance were conducted on each of the rating scale items. No main effects of THAMBS were significant regarding the rating scale items.

### 6.2. Effects of back-channelling on user behaviour and ratings

The second research question examined the effect of back-channelling on user behavior. Back-channelling in general led to few significant differences in user behaviour. Back-channelling fell just short of significance as a main effect on participant pauses, $F(2,72) = 3.07$, $p = .05$. Most pauses occurred in the back-channelling-*random* condition $M = .79$ ($SD = .66$), followed by back-channelling-*off* $M = .56$ ($SD = .58$), and then back-channelling-*on*, $M = .33$ ($SD = .70$). The contrast between back channelling-*on* and back-channelling-*random* conditions was significant, $F(2,67) = 6.14$, $p = .016$, Cohen's $d = .68$. There was no other main effect of back-channelling on user behavior, i.e., the number of comparisons made, attention, or eye contact, or on rating scale responses.

### 6.3. Interactions between factors and correlations between measures

The third research question considered interactions between THAMBS and back-channelling. There were no significant interactions between THAMBS and back-channelling on any of the behavioural dependent variables or rating scale responses.

Rating scale responses did not correlate with the video-based dependent variables, suggesting that the explicit subjective ratings are independent of participant behaviour in the experiment.

We also checked correlations between the video-based dependent variables (see Table 4). There were three significant correlations with $p < .001$. First, there was a significant negative correlation (−.44) between participants attracting Segie's attention and the number of comparisons that participants made: the more they attracted Segie's attention, the fewer comparisons they made. Second, there was a significant positive correlation (.39) between attracting attention and pauses: with more pauses made, the more participants attempted to re-engage or attract Segie's attention. Third, there was a significant positive correlation (.39) between eye contact and attracting attention with an association between more eye contact and more attempts to attract attention.

## 7. Discussion

An experiment investigated the influence of an active versus random attention model and back-channelling of a robotic "talking head" on user behavior. User behaviour included performance on a cognitive “robot comparison” task and behavioural (video-coded) and self-report ratings of engagement. Results were striking in the way that the user's behavior was influenced by the activation of the attention model with users working hard to engage and re-engage the robot's attention. For example, users gestured (waved their arms, clicked their fingers), spoke to (“Over here”, “C'mon Dude”), and made eye contact with the robot's avatar face significantly more often when the attention model—THAMBS—was *on* than when *random*. As users took time from the task for the interactive robot with THAMBS-*on* to re-orient, less time was available for the robot comparison task. Effects of back-channelling were independent of effects of the attention model. Back-channelling appeared seamless when active but was disruptive, i.e., elicited more user pauses, when the back-channelling activity was randomly timed. THAMBS and back-channelling had no effect on self-report ratings. In short, the presence of the attention model, and to a lesser extent back-channelling, led to a dynamical system of attention and engagement between the robot-mounted Thinking Head and the user.

The dependent variables or measures in the present experiment can be categorised as: behavioural indicators of engagement (user pauses, re-engaging Segie's attention, eye contact with Segie); self-report ratings of engagement; and task (robot comparison) performance. Each measure provides insight into the impact of the attention model and back-channelling on human–robot interaction. Of the behavioural indicators, there are effects of the salience of THAMBS, with users significantly more often engaging and re-engaging Segie's attention when THAMBS is on than when THAMBS operated in a random manner. This comparison shows that it is not just any movement, but attention-oriented, predictable movement that has an impact on user behavior. Participants waved their arms, moved to where Segie was fixating, clicked their fingers, or spoke to him—“Where are you going, Segie?”; “C'mon, Dude”. Such overt gestural and spoken communication reflects the users' Theory of Mind (Baron-Cohen et al., 1985; Premack and Woodruff, 1978) with respect to Segie and user assumption of the interactive robot possessing intelligence and agency. Theory of Mind refers to the ability to impute mental states to self and to others and to predict behavior on the basis of such states (Leslie, 1987). The present results align with those of Yamazaki et al. (2008) who demonstrated effect of robot head turn and gaze at interactionally significant points on human user non-verbal actions. The user behaviour that manifested in the present scenario demonstrated users' intention to guide or influence the robot's behavior. As Zhang et al. (2010) and Yu et al. (2012) have noted, humans are very sensitive to momentary multimodal behaviours generated by a virtual agent and humans adapt their behavior accordingly. Users in the present study also made significantly more eye contact with Segie when THAMBS was *on* compared with when THAMBS was *random*. Activation of the attention model in the robot appears to regulate user behaviour (Cassell et al., 1999; Salem et al., 2011).

Performance on the robot comparison task also reflects the effect of the attention model on users' engagement with the interactive robot with significantly fewer comparisons made when the attention model was on. Users' attention was drawn away from the comparison task to interact with, or attend to, the interactive robot. In practical terms, when the robot is moving, the time the participants are paying attention to it is taken from the time they spent explaining more differences between the robots. This result is key to understanding the complex effects and ramifications of a socially engaging robot. While poorer performance on the cognitive task in the present experiment was not life threatening, there would be situations where distraction of the user from their task by an agent is less than ideal.

Back-channelling also influenced user behavior but in different ways from that of the attention model. Back-channelling appears distinct from “attention” cues in the present task. Movement associated with the attention model being active was potent but did not interact with language. Specifically, users tended to pause more when back-channelling was random than when it was on. This suggests that back-channelling did not interrupt the interaction but was well integrated with the task and robot–user interaction.

Our results lead us to conclude that more “believable” or realistic attention-display behaviour from the robot elicits a commensurately higher expectation of social behaviour by the robot from the human. The robot displaying (through its behaviour) that it is capable of the social cues can be seen as leading to an expectation from the human subjects that they reciprocate corresponding social behaviours whenever appropriate. In fact, this behavior can even be to the detriment of actual task performance: subjects spend more time trying to engage the robot when the attention model is activated and ignoring the robot more when it is not, leading to improved task performance in the latter case. This is, in our opinion, an important consideration in the design of robot-interaction behavior, depending on whether a robot's purpose is intended to be perfunctorily task-oriented or more social in nature. While not unrelated to the Uncanny Valley effect (Mori et al., 2012), which hypothesises that increasing realism in a virtual avatar eventually leads to revulsion from the user, the described effects are different: subjects are not so much repulsed by the more realistic social behaviours but do seem to set higher expectations of the robot because of them.

Future investigations will develop the interactivity of the setting using a turn-taking game, with greater interaction time and ecological validity, and will examine the role of more sophisticated language use and its impact on attention within the context of a conversation. While needing to maintain control to optimise generalisability of results (e.g., Banaji and Crowder, 1989) ecological validity of the scenario and human–agent interaction can also be augmented.

## Acknowledgements

This research was supported by the Thinking Head project, a Special Initiative scheme of the Australian Research Council and the National Health and Medical Research Council (TS0669874).

We thank the anonymous reviewers for extremely valuable feedback, insights, and suggestions.

## References

Argyle, M., Cook, M., 1976. Gaze and Mutual Gaze. Cambridge University Press, New York.

Bailly, G., Raidt, S., Elisei, F., 2010. Gaze, conversational agents and face-to-face communication. Speech Commun. 52 (6), 598–612.

Baron-Cohen, S., Leslie, A.M., Frith, U., 1985. Does the autistic child have a “theory of mind”? Cognition 21, 37–46.

Banaji, M.R., Crowder, R.G., 1989. The bankruptcy of everyday memory. Am. Psychol. 44 (9), 1185–1193.

Bavelas, J.B., Coates, L., Johnson, T., 2000. Listeners and co-narrators. J. Personal. Soc. Psychol. 79 (941), 952.

Bee, N., Wagner, J., André, E., Vogt, T., Charles, F., Pizzi, D., Cavazza, M., 2010. Discovering eye gaze behavior during human–agent conversation in an interactive storytelling application. In: Proceedings of the International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI '10). ACM, New York, NY, USA.

Bosse, T.V., Maanen, P.-P., Treur, J., 2006. A cognitive model for visual attention and its application. In: Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IEEE Computer Society.

Breazeal, C., 2002. Designing Sociable Robots. The MIT Press.

Breazeal, C., Scassellati, B., 1999. Context-dependent attention system for a social robot. In: Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, pp. 1146–1153.

Brunner, L.J., 1979. Smiles can be back channels. J. Personal. Soc. Psychol. 37 (5), 728–734.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhj, H., et al., 1999. Embodiment in conversational interfaces: Rea. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: the CHI is the Limit, Pittsburgh, pp. 520–527.

Colburn, R.A., Cohen, M.F., Drucker, S.M., 2000. The Role of Eye Gaze in Avatar Mediated Conversational Interfaces. Microsoft Research Technical Report: MSR-TR-2000-81, July.

Fong, T., Nourbakhsh, I., Dautenhahn, K., 2003. A survey of socially interactive robots. Robot. Autonomous Syst. 42, 143–166.

Fujie, S., Fukushima, K., Kobayashi, T., 2005. Back-channel feedback generation using linguistic and nonlinguistic information and its application to spoken dialogue system. In: Proceedings of INTERSPEECH, Lisbon, Portugal.

Goldin-Meadow, S., 1999. The role of gesture in communication and thinking. Trends Cogn. Sci. 3 (11), 419–429.

Goodrich, M.A., Schultz, A.C., 2007. Human-robot interaction: a survey. Found. Trends Hum.-Comput. Interact. 1 (3), 203–275.

Herath, D., Kroos, C., Stevens, C., Cavedon, L., Premaratn, P., 2010. Thinking Head: towards human centred robotics. In: Proceedings of the 11th International Conference on Control, Automation, Robotics, and Vision (ICARCV), Singapore, pp. 2042–2047.

Hess, L.J., Johnston, J.R., 1988. Acquisition of back channel listener responses to adequate messages. Discourse Process. 11 (3), 319–335.
Hiroji, I., 2006. Learners' perception and interpretation of contextualization cues in spontaneous Japanese conversation: back-channel cue Uun. J. Pragmat. 38 (11), 1943–1981.
Imai, M., Ono, T., Ishiguro, H., 2003. Physical relation and expression: Joint attention for human–robot interaction. IEEE Trans. Ind. Electron. 50 (4), 636–643.
Jonsdottir, G.R., Gratch, J., Fast, E., Thórisson, K.R. , 2007. Fluid semantic back-channel feedback in dialogue: challenges and progress. In: Proceedings of the Intelligent Virtual Agents, Paris, LNAI 4722, pp. 154–160.
Kim, Y., Hill, R.W., Traum, D.R., 2005. A computational model of dynamic perceptual attention for virtual humans. In: Proceedings of the 14th Conference on Behavior Representation in Modeling and Simulation (BRIMS), Universal City, CA.
Kopp, S., Allwood, J., Grammer, K., Ahlsen, E., Stocksmeier, T., 2008. Modeling embodied feedback with virtual humans. In: Proceedings of the Embodied Communication in Humans and Machines, 2nd ZiF Research Group International Conference on Modeling Communication with Robots and Virtual Humans, Springer, LNCS 4930, pp. 18–37.
Kopp, S., Stocksmeier, T., Gibbon, D., 2007. Incremental multimodal feedback for conversational agents. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (Eds.), Intelligent Virtual Agents, vol. 4722. Springer, Berlin/Heidelberg, pp. 139–146.
Kroos, C., Herath, D.C., Stelarc, 2011. From robot arm to intentional agent: the articulated head. In: Goto, S. (Ed.), Robot Arms. Advances in Robotics, Automation and Control. InTech, pp. 215–240.
Kroos, C., Herath, D.C., Stelarc, 2012. Evoking agency: Attention model and behaviour control in a robotic art installation. Leonardo 45 (5), 133–161.
Langton, S.R.H., Bruce, V., 2000. You must see the point: automatic processing of cues to the direction of social attention. J.Exp. Psychol.: Hum. Percept. Perform. 26 (2), 747–757.
Langton, S.R.H., Watt, R.J., Bruce, V., 2000. Do the eyes have it? Cues to the direction of social attention. Trends Cogn. Sci. 4 (2), 50–59.
Leslie, A.M., 1987. Pretense and representation: the origins of "Theory of Mind. Psychol. Rev. 94 (4), 412–426.
McNeill, D., 2005. Gesture and Thought. University of Chicago Press, Chicago, IL.
Mori, M., MacDorman, K.F., Kageki, N., 2012. The uncanny valley [from the field]. IEEE Robot. Autom. Mag. 19 (2), 98–100.
Mutlu, B., Forlizzi, J., Hodgins, J., 2006. A story-telling robot: modelling and evaluation of human-like gaze behaviour. In: Proceedings of the 6th IEEE-RAS International Conference on Humanoid Robots, Genoa, Italy.
Peters, R.J., Itti, L., 2006. Computational mechanisms for gaze direction in interactive visual environments. In: Proceedings of the Symposium on Eye Tracking Research & Applications, San Diego, California, pp. 27–32.
Poggi, I., 2005. Backchannel: from humans to embodied agents. In: Proceedings of the Conversational Informatics for Supporting Social Intelligence and Interaction—Situational and Environmental Information Enforcing Involvement in Conversation Workshop in AISB'05, University of Hertfordshire, Hatfield, England.
Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? Behav. Brain Sci. 4, 515–526.
Rajan, S., Craig, S.D., Gholson, B., Person, N.K., Graesser, A.C., 2001. Autotutor: incorporating back-channel feedback and other human-like conversational behaviors into an intelligent tutoring system. Int. J. Speech Technol. 4 (2), 117–126.
Salem, M., Rohlfing, K., Kopp, S., Joublin, F., 2011. A friendly gesture: investigating the effect of multi-modal robot behavior in human-robot interaction. In: Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication, Atlanta, Georgia, pp. 247–252.
Starkey, D., 1972. Some signals and rules for taking speaking turns in conversations. J. Pers. Soc. Psychol. 23 (2), 283–292.
Sun, Y., Fishe, B., Wang, H., Gomes, M., 2008. A computer vision model for visual-object-based attention and eye movements. Comput. Vis. Image Underst. 112, 126–142.
Ward, N., Tsukahara, W., 2000. Prosodic features which cue back-channel responses in English and Japanese. J. Pragmat. 23, 1177–1207.
Yamazaki, A., Yamazaki, K., Kuno, Y., Burdelski, M., Kawashima, M., Kuzuoka, H., 2008. Precision timing in human–robot interaction: coordination of head movement and utterance. In: Proceedings of the CHI'08, Florence, Italy, pp. 31–140.
Yu, C., Schermerhorn, P., Scheutz, M., 2012. Adaptive eye gaze patterns in interactions with human and artificial agents. ACM Trans. Interact.Intell. Syst. 1 (2) 1-13:25.
Zhang, H., Fricker, D., Yu, C., 2010. A multimodal real-time platform for studying human–avatar interactions. In: Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA 2010), Philadelphia, PA.