Estimating Human Interest and Attention via Gaze Analysis

Heather Knight
Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15224
Email: hknight@cs.cmu.edu

Reid Simmons
Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15224
Email: reids@cs.cmu.edu

***Abstract*—In this paper we analyze joint attention between a robot that presents features of its surroundings and its human audience. In a statistical analysis of hand-coded video data, we find that the robot's physical indications lead to a greater attentional coherence between robot and humans than do its verbal indications. We also find that aspects of how the tour group participants look at robot-indicated objects, including when they look and how long they look, can provide statistically significant correlations with their self-reported engagement scores of the presentations. Higher engagement would suggest a greater degree of interest in, and attention to, the material presented. These findings will seed future gaze tracking systems that will enable robots to estimate listeners' state. By tracking audience gaze, our goal is to enable robots to cater the type of content and manner of its presentation to the preferences or educational goals of a particular crowd, e.g. in a tour guide, classroom or entertainment setting.**

I. Introduction

Joint attention is a powerful tool for conversation partners to establish repertoire and increase the fluency of communication [6], [8], [9]. Research shows that listeners can use speaker gaze and gestural indications to identify the intended subject matter before the linguistic point of disambiguation (i.e., the point in a description where possible interpretations can be eliminated and the intended sentence meaning is clear) [4].

In the study presented here, a stationary Nao robot acts as a tour guide (see Fig. 1) presenting locational information about one of three spaces in one of three presentation styles. Via statistical analysis of the hand-coded video data, we evaluate three hypotheses: (1) whether there is coherence between listener gaze and the location the robot is presenting, (2) whether overlaying robot deictic gestures on top of verbal discourse further directs the listener gaze to an indicated scene or object, and (3) whether there is a relationship between the audience's gaze responses to robot deictic gestures and their overall engagement. A deictic gesture is an indicatory gesture, such as pointing, but may also be achieved with arm position, body orientation or an indicatory nod. We use participants' self-reported engagement scores as ground truth.

Our evaluation of robot indicatory strategies finds that physical gestures, namely, the robot gesturing toward a location with its whole arm, led to the greatest number of listeners looking in that direction, as compared to verbal references only. This suggests that the robot's physicality, in both gaze

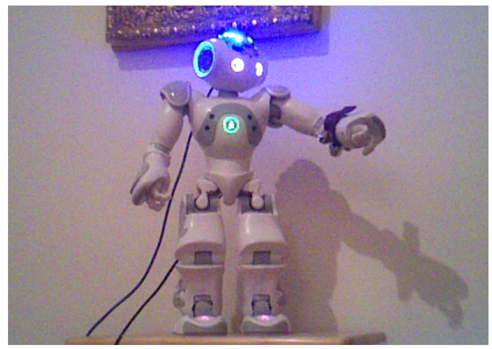


Fig. 1. Sample image of Nao robot indicating location during a presentation.

Fig. 2. Tour group responding to robot deictic indication of the ceiling during a presentation of the fire emergency system. The robot's hand is visible in the bottom left corner of the image.

and gesture, plays an important role in the overall coordination of information. We also find trends indicating features of how individual listeners look that correlate with their overall engagement. Looking more quickly toward the indicated direction, gazing for longer, and glancing back more frequently are all weakly associated with higher engagement. The first of these two features demonstrate statistical significance at the group level, i.e., aggregating these features over the group allows us to predict average group engagement at a better than chance level. This suggests that a robot interpreting cues from a coherent group might be more reliable than one looking at individuals only, as human behavior is complex and noisy.

II. Related Work

Relevant prior research includes human gaze tracking analysis, research into how robot gaze impacts conversational partners, and related domain research in which a robot tour guide or presenter was used as a domain to evaluate human-robot interaction principles. The central work of this paper will build on these investigations, providing a useful new set of features for interpreting participant response nonverbally via gaze analysis in the temporal vicinity of robot deictic gestures.

A. Human gaze in the context of robots

A robot that tracks where its interaction partner is looking might better understand their state. In [20], a robot tour guide scans the group after asking a question to decide whom it might ask for the answer. If someone is not making eye-contact, they probably do not want to be called on, thus the robot will decide whom to single out after finding someone who appears to be meeting its gaze. In [17], the robot that acknowledges listener head nods by nodding back and/or accepting a nod as an affirmation, increases participants overall use of head nodding.

Human gaze can also help a robot understand their focus of attention and intention. In [15], a robot uses gaze estimates to help a human to assemble legos, e.g. when a person pauses for a while, the robot might infer which piece they are currently assessing for fits and offer sample connecting pieces, or add a piece that the subject is eyeing, when it is clear their hands are full.

Finally, estimating human gaze during an interaction helps guide a conversation partner's understanding of interest, attention and social role [6].

B. Robot gaze

Many studies show that humans automatically read and interpret non-verbal social behaviors from a robot and that robot gaze influences the behavior of human interaction partners, e.g. indicating social roles, directing attention or influencing participation [2], [5], [10], [18], [14].

In [10], Robovie uses gaze cues to signal to certain people that they are either current conversation partner(s), seen but next in line, or not part of any interaction. The work evaluated the success of robot-indicated role-assignment by who would answer the robot's questions, finding that subjects treated as conversation partners were most likely to respond to the robot. In finer detail, the robot gaze might establish the communication target(s), maintain indication of current speakers, and guide turn-taking. The work, however, did not explicitly evaluate joint gaze toward objects.

The work in [18] demonstrated that a robot's gaze helps direct where people look in a scene, improving (or ambiguating) their understanding of the robot's utterances, depending on whether the gaze was congruent (or not) with the speech. The robot's gaze was used to indicate which object it was referring to, and aided in the fluency of the communication.

Additional work shows that a robot can indicate its engagement with its gesture pattern [16]. In interactions with gestures, people direct their attention to the robot more often, also finding the robot behavior more appropriate than in cases where the robot expresses no gestures.

C. Related domains

Though [3], [7], [11] discuss tour guide robots that can play games, button-in topic selection or have simple dialog and/or dance, much of the early work in robot tour guides involved strategies to navigate safely through museums [1], [13].

The work in [12] reported that a robot tour guide that moves backwards results in more bystanders pausing to overhear the robot's presentation, some of whom will join the robot's audience. They hypothesized that the robot's body orientation made it more socially acceptable for new people to look on, as there was a wider view area for the robot to include new people in its visible range while backwards-facing, rather than glancing sideways in the forward-facing condition.

The ethnographic analysis in [19] showed that both tour guides and visitors naturally spend significant amounts of time looking at objects of attention during an interaction. Their robot tour guide presented historic material about paintings in an art museum, using involvement questions to engage possible listeners, such as asking if they wanted to hear something about the artwork. Based on estimated head pose toward a painting, the robot would classify them as 'highly interested' and would tell them more.

Although a nice starting point, this binary system lacks the nuance of complex gaze, assessing only if it exists, and does not allow for an engagement analysis across multiple visitors. It might also lead to cases where the overall tour stops unnecessarily due to oversimplifications in the social model, although they do use additional pointing gestures and small repetitions of utterance to try to regain participant attention. We think this would make a nice additional feature of our future crowd-adaptive robot tour guide.

To design and evaluate their social design, many of the above experiments use hand-labeled video data, whether to estimate natural gaze behaviors or to interpret human responses to a robot orientation or behavior. We build on such techniques to analyze our own study data.

III. PROCEDURE

In this study, we assess human gaze in response to a Nao robot tour guide that is presenting features of its environs. Specifically, we evaluate our three hypotheses: (1) whether there is coherence between listener gaze and the location the robot is presenting, (2) whether overlaying robot deictic gestures on top of verbal topics further directs the listener gaze to an indicated scene or object, and (3) whether there is a relationship between human gaze response to robot deictic gestures and their overall engagement.

The experiment takes place in the modern and uniquely constructed Gates Building on the Carnegie Mellon campus, with many interesting architectural features, installations, and views. There were thirty-four participants overall, split into nine tour groups of three to five people each, including a ethnographically representative range of friends, strangers and families, i.e., similar to the makeup of typical campus tours. The participants were unfamiliar with most the material the

Fig. 3. Robot Tour Locations: a) "The Helix" location is atop a large spiraling walkway, its presentation discusses architectural features and the displayed software art Electric Sheep, b) "Green Garden" discusses the rooftop garden out the window and sustainable design of the building, c) "Ace Atrium" discusses the unusual fire safety mechanisms in this student lounge and statistics from the CMU Robotics Institute, which is on view out the window.

robot presented. Further participant details are outlined in Table I.

A. Study Design

In order to evaluate our hypotheses relating audience gaze to attention and engagement, there must be a wide range of group reactions. To enable this, we varied topics by location (helix, garden, atrium, see Fig. 3) and robot speech patterns and motions by presentation style (friendly, neutral, sarcastic) in nine Latin Square combinations. This was an intra-participant study in which each tour group saw three tour guide presentations that spanned all locations and all robot presentation styles, but not every permutation therein, as represented in Table II.

All participants completed brief surveys after each presentation and a final survey that rated the different presentations comparatively. In pilot studies, both friendly and sarcastic presentations of the same factual material resulted in higher self-reported engagement scores as compared to neutral, which we intended to use to help evaluate whether we could use participant gaze responses to automatically identify engagement level.

Audiences were video-taped from the robot's perspective using a Kinect sensor. While the work reported here uses the RGB images to code participant head pose, we intend to eventually include the Nao as the communicatory social presence standing atop a non-anthropomorphic robotic base equipped with various sensors, hence the applicability of using a Kinect. In our current study, however, the robot stood on top of a tall table.

We used a subset of the overall dataset for the current analysis. A single coder labeled participant head pose for nine of the twenty-seven three-minute videos. This subset covers thirteen participants across the 15-frames-a-second footage (about 24300 frames total). We also label the timings of robot topics (based on utterances) and deictic indications (based on robot gesture), as represented in Fig. 4. The correspondence between robot and audience behaviors enables us to evaluate our original hypothesizes.

B. Locations and Presentation Styles

Each group experienced presentations in three locations in the building (Fig. 3): a student lounge called "Ace Atrium", a workspace overlooking a garden, aka, "Green Garden", and a major entrance to the building "The Helix". Each presentation references three or four major features or topics related to the space, e.g., architectural features, artwork, the view, or an unusual fire-safety system.

To illustrate the coherence of factual material across the robot presentation styles, we present dialog excerpts from the "Green Garden" presentations:

Green Neutral:

Hello. I will begin my presentation now. Here you get a clear view of one of the six green garden roofs in the building. This garden uses real soil and dirt samples. Most of the other roofs sustain their plants in special hydroponic tray systems that don't use real soil. That's the difference.

Green Sarcastic:

Another tour group? Man! Let's make this quick! They tell me that that plot of dirt out the window is one of the six green garden roofs in the building. You might be thinking, so what! Me too. But, this garden uses real soil and dirt samples. The other roofs sustain their plants in special hydroponic tray systems that don't use real soil.

Green Pleasant:

Welcome everybody! What a great building, don't you think? Here you get a great view of one of the six green garden roofs in the building! So cool, right? This garden actually uses real soil and dirt samples. Most of the other roofs sustain their plants in special hydroponic tray systems, also pretty neat!

We designed the robot speech and gestures in conjunction with CMU Professor of Acting Matthew Gray. Though each robot presentation contains the same factual material and sequence of deictic gestures for each location, its base pose, conversational gesture heights and choice of adjectives changed to be coherent with its given presentation style.

The robot references topics verbally 5-8 times and physically indicates their locations with head and limb deictic gestures 2-4 times per presentation.

C. The User Study

After completing demographic surveys and introducing participants to the activity, the study conductor led them to the first of three robots. This robot greeted them and talked about features of the room, view or objects in the space in one of the three presentation styles. Led to the second space and

TABLE I
PARTICIPANT DEMOGRAPHICS

Number	Gender	Age	Tech Savvy	Comfortable with Robots
34	13 Female / 21 Male	14-61 (mean 27, std 13)	92%	5 No / 3 Sort-of / 26 Yes

TABLE II
LATIN SQUARE OF LOCATION-PRESENTATION STYLE PAIRINGS, INDICATING TOUR GROUP NUMBER - TRIAL NUMBER

	The Helix	Green Garden	Ace Atrium
Please	1-1	3-3	2-2
Neutral	2-3	1-2	3-1
Unplease	3-2	2-1	1-3

robot, the exercise repeated, this time with different topics and personality, and again for the third location.

Table II summarizes the tour group order for the data analyzed in this report. For example, Tour Group 1 first saw the friendly robot at The Helix location, then they saw the neutral robot at the Green Garden location and finished with the sarcastic robot at the Ace Atrium. Thus, all tour groups saw all presentation styles and all locations, but in different combinations. After three tour groups, the full possible table of location-presentation style pairings is explored. Groups 4-6 and 7-9 repeat the ordering seen for Groups 1-3.

At the completion of each robot presentation, the participants completed a short survey. The questions used to estimate engagement include two 7-point scales. In the first, participants rated "I found the robot's presentation entertaining" from Strongly Disagree to Strongly Agree, and in the second, they rate the robot itself from Boring to Engaging.

D. Data Analysis

After data collection, we labeled participant gaze in a subset of the data that fully explores one Latin Square (see Table II), using the RGB video data to code viewer gaze statistics.

1) Coding: The nine videos with 3-5 audience members each were hand-labeled by the lead author to indicate the head pose of each participant. Directions include up, down, camera-left, camera-right, straight ahead, behind. We selected this coarse labeling system for two reasons: the tractability of future machine vision implementations, and, more importantly, the level of detail was sufficient to disambiguate possible objects of attention as prescribed by the locational features indicated in the robot's presentations.

These gaze directions were next matched to objects-of-attention in the scene, e.g., a scene indicated out the window might be on the right, and intra-participant gaze involves spatially matching two participants looking toward each other. The objects-of-attention presented during the robot presentation were chosen to be non-overlapping spatially to simplify analysis.

We also established timing and topic labels for the nine robot scripts,including tracks for overall topic, robot gaze and deictic gestures (see Fig. 4 for sample). In our study, deictic

Notice those things that look like doors? (66-118)

Well, they are doors.

The tall rectangles of glass are motorized doors that can open upon command.

This is why...

The fire code demands that, in case of fire, atriums must have a way of removing smoke.

So they have installed big fans on the top of that **wall.

See the gray metal up there?

** (574-832) The reasoning is a simple calculation, if you pull lots of air out of the room, you need to let more air in. So when the fan goes on, the doors automatically *open.*(1109-1132)

And that's that.

Fig. 4. Excerpt of from robot presentation dialog, with labels for deictic gestures (underlined), and the corresponding video frames over which they occur.

refers to large arm gestures referencing a physical location, an example of which is in Fig. 1. The timing and deictic labels allow us to match the robot's current focus of attention to the audience gaze, and understand whether talking about a location, versus looking at the location, versus pointing at the location, is most impactful in guiding listeners' attention.

2) Analysis: To evaluate our first and second hypotheses, we first isolate participant gaze directions, and perform a binary matching over when the participant is gazing at an object of interest and when the robot has that area as its current focus. We use a series of contingency analyses to indicate the level of coherence between robot topic and audience attention, and whether the presence of a physical gesture (robot gesture and head pose) in addition to verbal indications increases the correlation. We use a t-test to establish the statistical significance of our findings.

To evaluate our third hypothesis, we combine statistics from all gaze directions to identify higher-level gaze-based features we could use to classify audience engagement using an ANOVA analysis. For example, if participants look toward an area that the robot is indicating, we assess temporal features of how they look, including the timestamp of their first look, how long they spend looking, the timestamp when they stop looking, their overall gaze shifts, and their total number of glance backs following the indication of the location.

IV. RESULTS

The default gaze direction in all groups is toward the robot, as it is the one talking. Thus, participants at all engagement levels spend large portions of time gazing at the robot. Our 13-person data subset showed that participants spent more than 70% of the tours gazing at the robot.

A. Coherence between robot topic and audience gaze

To evaluate our first hypothesis, that there would be a coherence between audience gaze and the location the robot is discussing, we performed a contingency analysis between the labeled topic of the robot utterance and audience head pose (see Fig. 5), which is consistent with our other data. The results

TABLE III
THE INFLUENCE OF ROBOT DEICTIC GESTURE ON PARTICIPANT GAZE. DATA FROM GREEN GARDEN-NEUTRAL CONDITION, GROUP 2. SIX DEICTIC GESTURE EVENTS FOR FOUR PARTICIPANTS, THUS N=24. * INDICATES STATISTICAL SIGNIFICANCE, ** INDICATES VERY STRONG STATISTICAL SIGNIFICANCE.

	Participant 1	Participant 2	Participant 3	Participant 4
p-test cond1	0.0156*	0.1317	.0197*	less than .0001**
p-test cond2	0.0014*	0.0044*	.0002**	.0001**

strongly indicate that participants are looking at locations the robot is discussing, with 91% of upward gaze occurring while the robot is discussing a feature on the ceiling and 99.2% of the participant gaze left (towards the window) during utterances about features and buildings on view out the window. Perhaps such analysis could help future robot tour guides automatically detect who is a participant versus bystander.

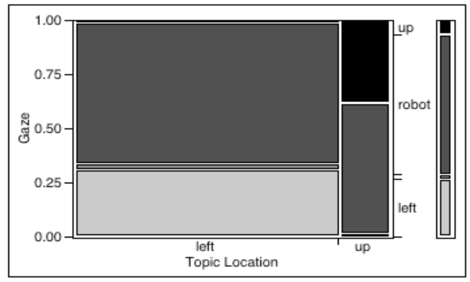
This does not imply, however, that audience gaze is mostly up while the robot is speaking about an upward feature. Instead 61% of the time when the robot was presenting an upward feature was spent gazing at the robot and 38% at the feature. Similarly, during the leftward topics, the audience spent 64% of their time looking at the robot and 31% gazing left. This result seems intuitive, as long as the audience is paying attention and attuned to the robot, but also provides us a benchmark for the comparative influence of gesture (the second hypothesis).

B. Coherence between robot gesture and audience gaze

To evaluate our second hypothesis, that adding a robot deictic gesture to its topic indication better directs audience gaze to the indicated scene or object, we perform a second contingency analysis. In Table III, we present a representative data sample. In "cond1," we do a p-test to evaluate the influence of specific robot deictic gesture (toward feature vs. no gesture) to specific audience gaze directions (up, down, right, left, back, robot). In "cond2," we evaluate the correspondence of any robot deictic gesture (present or not present) with audience gaze in any non-robot direction (toward robot or toward anything else). Deictic gestures show statistical significance for almost all cases.

Doing a third contingency analysis that isolates the timestamps where a deictic gesture is present results in an even stronger correlation. The co-variance measure between robot gesture and participant gaze in the Ace Atrium condition is 0.248, so they are independent, but a contingency analysis shows that for N=140, 100% of the participant ceiling gaze (during any deictic gesture) occurs during robot up gestures and 100% of the window gaze (during any deictic gesture) occurs during robot gestures toward the window.

As compared to the robot utterances case in Fig. 6, participants spend 57% of the time looking upwards during a robot 'up' gesture, whereas they had only spent 38% of the time looking up when the feature in that location was discussed verbally. Similarly, they spend 50% of their time looking out the window during robot deictic gestures toward the



Gaze

Topic Location	Count Total % Col % Row %	down	left	right	robot	up	
	left	1 0.21 100.00 0.25	128 26.45 99.22 31.37	9 1.86 100.00 2.21	267 55.17 85.30 65.44	3 0.62 9.38 0.74	408 84.30
	up	0 0.00 0.00 0.00	1 0.21 0.78 1.32	0 0.00 0.00 0.00	46 9.50 14.70 60.53	29 5.99 90.63 38.16	76 15.70
		1 0.21	129 26.65	9 1.86	313 64.67	32 6.61	484

Fig. 5. Contingency Analysis by Topic Location for Ace Atrium, neutral presentation style, where topic indicates the subject of robot utterances. Sampled any time there was a participant gaze change over the three minute presentation (N=484) and Prob$\geq$ ChiSq less than .0001, thus, there is a very significant effect of robot topic on participant gaze direction.

window, as compared to 31% during topical robot utterances, demonstrating that robot deictic gestures do in fact have a large role in influencing participant gaze and attention, one that has more impact than utterances alone.

C. Predicting listener engagement from their gaze features

Finally, we evaluate our third hypothesis, that the listener gaze response to robot deictic gestures will allow us to predict their overall engagement.

We first note that we achieved a good spread of engagement scores across the different robot presentations (possible scores were from 1 for boring to 7 for engaging), as presented in Fig. 6, which validates our method of utilizing different robot presentation styles to prompt a range of participant responses. The text box for the figure shows that some participants may have been more or less inclined to be engaged to begin with, but robot presentation style had more than three times the impact on their engagement level than who they were (p=0.0071 as opposed to p=0.0271).

In examining the relationship between calculated features of participant gaze (time until first look, time spent looking, how long until they stop looking, total gaze shifts during gesture, total glance backs to topic) for each robot gesture and predicting participant engagement for a single deictic gesture, we obtain mixed results (Table IV). There are trends that "first look," "time spent looking," and "glance backs" would be good indicators for estimating listener engagement. This signal strengthens at the group level, as there was intra-group coherence. Averaging a tour group's gaze features gives a statistically significant prediction of that group's overall level of engagement for "first look" and "time spent looking."

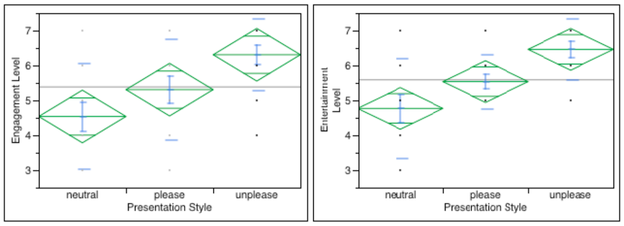


Fig. 6. Analysis of 13 participants across three scenes (N=39) shows the very significant influence of presentation style on participants' self-reported engagement (p=0.007) and entertainment scores (p=0.001). The influence of the participant identity was significant for engagement (p=0.027), but not significantly above chance for entertainment (p=0.195).

TABLE IV
INFLUENCE OF GAZE FEATURES ON THE PREDICTION OF ENGAGEMENT, FOR INDIVIDUALS AND GROUPS (13 INDIVIDUALS, 3 GROUPS), GIVEN A FIVE DEICTIC GESTURES ACROSS ALL PRESENTATION STYLES IN THE ACE ATRIUM (N=65 FOR INDIVIDUAL DATA, N=15 FOR GROUP DATA). POSITIVE OR NEGATIVE CORRELATION INDICATED BY (+) OR (-).

	when first look (-)	time spent looking (+)	when stop looking (-)	total shifts (+)	glance backs (+)
Individ.	0.1500	0.1125	0.8637	0.4699	0.0963
Group	0.0085*	0.0014*	0.2056	0.3816	0.1089

When you combine these trends across the 5-6 deictic gestures in each presentation, the evidence becomes stronger. It is possible that machine learning techniques could further strengthen the signal.

V. CONCLUSION

This paper seeks to find a relationship between robot topic and participant gaze that would allow a robot tour guide to assess audience attention and engagement level and attune to its audience. We also evaluate the influence of robot deictic (pointing) gestures on participant attention, as compared to verbal topic only.

We demonstrate the robot utterances have a significant influence on where the audience might look. Although they spent most of the time looking at the robot, when they did look elsewhere, it was most often (over 90% of the time) contingent with the current robot topic. We also demonstrate that a robot's deictic gestures are more effective than utterance alone at triggering audience gaze, which implies they are better at eliciting listener attention. Finally, we display how higher level features of participant gaze in the temporal vicinity of each deictic gesture could help a robot predict the current engagement levels of a crowd.

By tracking the gaze responses of participants in response to a robot's deictic gestures, we gain an indication of participant interest and attention to the subject matter or objects presented. The techniques we will build based on these learnings expand the social intelligence capabilities of robots interacting with groups of people, and should equally cross-apply to classroom or entertainment settings.

ACKNOWLEDGMENT

The authors would like to thank the National Science Foundation and their Graduate Research Fellowships Program. Special thanks to Professor of Acting Matthew Gray who helped design robot gestures coherent with the three robot presentation styles.

REFERENCES

[1] M. Bennewitz, F. Faber, D. Joho, M. Schreiber, and S. Behnke, Towards a humanoid museum guide robot that interacts with multiple persons. In Proc. HUMANOIDS 2005, pp.418-423.
[2] D. Bohus, E. Horvitz, Facilitating multiparty dialog with gaze, gesture, and speech, Proc. ICMI-MLMI International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, 2010.
[3] K. Gunhee et al, The Autonomous Tour-Guide Robot Jinny, in Proc. IROS, 2004.
[4] J. Hanna and S. Brennan, Speakers eye gaze disambiguates referring expressions early during face-to-face conversation, Journal of Memory and Language, 2007, 57:596615.
[5] S. Hinte, M. Lohse, The Function of Off-Gaze in Human-Robot Interaction, RO-MAN 20th IEEE Intl Symposium on Robot and Human Interactive Communication, 2011.
[6] C. Huang, A. Thomaz, Effects of Responding to, Initiating and Ensuring Joint Attention in Human-Robot Interaction, in Proc. RO-MAN 2011.
[7] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, An affective guide robot in a shopping mall, In Proceedings of the 4th ACM/IEEE Proc. of Intl Conf. on Human-Robot Interaction, 2009, pp. 173-180.
[8] F. Kaplan, V. Hafner, The Challenges of Joint Attention, Proc. of the Fourth Intl Workshop on Epigenetic Robotics, 2004, pp. 67-74.
[9] C.Moore, P. Dunham, editor. Joint Attention Its Origins and Role in Development, LEA, 1995.
[10] B. Mutlu et al Footing In Human-Robot Conversations: How Robots Might Shape Participant Roles Using Gaze Cues, In Proc. Intl Conf on Human-Robot Interaction, 2009, pp. 61-68.
[11] I.R. Nourbakhsh, C. Kunz, T. Willeke, The mobot museum robot installations: a five year experiment, In Proc. IROS 2003.
[12] M. Shiomi, M. Kanda, T. Ishiguro, H. Hagita, A larger audience, please! Encouraging people to listen to a guide robot, In Proc. Intl Conf. on Human-Robot Interaction, 2010.
[13] S. Thrun, et al. MINERVA: a second-generation museum tour-guide robot, In Proc. Intl Conf. on Robotics and Automation, 1999.
[14] K. Sadazuka, Y. Kuno, M. Kawashima, K. Yamazaki, Museum Guide Robot with Effective Head Gestures, In Proc. Intl Conf. on Control, Automation and Systems, Oct. 17-20, 2007.
[15] K. Sakita, K. Ogawara, S. Murakami, K. Kawamura, K. Ikeuchi, "Flexible cooperation between human and robot by interpreting human intention from gaze information," in Proc. Intl Conf. on Intelligent Robots and Systems, vol.1, no., pp. 846- 851 vol.1, 28 Sept.-2 Oct. 2004.
[16] C.L. Sidner, C. Lee, C.D. Kidd, C. Rich, Explorations in engagement for humans and robots, Artificial Intelligence, 2005, 166(1), pp. 140-164.
[17] C. L. Sidner, C. Lee, L. Morency, C. Forlines, The effect of head-nod recognition in human-robot conversation, In Proc. Intl Conf. on Human-Robot Interaction, 2006.
[18] M. Staudte, M. W. Crocker, Visual attention in spoken human-robot interaction, In Proc. ntl Conf. on Human-Robot Interaction, 2009 pp. 77-84.
[19] K. Yamazaki, A. Yamazaki, M. Okada, Y. Kuno, Y. Kobayashi, Y. Hoshi, K. Pitsch, P. Luff, D. Lehn, C. Heath, Revealing Gauguin: Engaging Visitors in Robot Guides Explanation in an Art Museum, New Media Experiences 2 April 8th, 2009.
[20] Y. Kobayashi, T. Shibata, Y. Hoshi, Y. Kuno, M. Okada, K. Yamazaki. Choosing answerers by observing gaze responses for museum guide robots, In Proc. Intl Conf. on Human-Robot Interaction, 2010, pp. 109-110.