Sensing and Handling Engagement Dynamics in Human-Robot Interaction Involving Peripheral Computing Devices

Mingfei Sun
Hong Kong University of Science and Technology
Hong Kong
mingfei.sun@ust.hk

Zhenjie Zhao
Hong Kong University of Science and Technology
Hong Kong
zzhaoao@cse.ust.hk

Xiaojuan Ma
Hong Kong University of Science and Technology
Hong Kong
mxj@cse.ust.hk

ABSTRACT
When human partners attend to peripheral computing devices while interacting with conversational robots, the inability of the robots to determine the actual engagement level of the human partners after gaze shift may cause communication breakdown. In this paper, we propose a real-time perception model for robots to estimate human partners' engagement dynamics, and investigate different robot behavior strategies to handle ambiguities in humans' status and ensure the flow of the conversation. In particular, we define four novel types of engagement status and propose a real-time engagement inference model that weighs humans' social signals dynamically according to the involvement of the computing devices. We further design two robot behavior strategies (*explicit* and *implicit*) to help resolve uncertainties in engagement inference and mitigate the impact of uncoupling, based on an annotated human-human interaction video corpus. We conducted a within-subject experiment to assess the efficacy and usefulness of the proposed engagement inference model and behavior strategies. Results show that robots with our engagement model can deliver better service and smoother conversations as an assistant, and people find the implicit strategy more polite and appropriate.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces - interaction styles.: H.1.2 Models and Principles: User/Machine Systems - human factors. I.2.9 Robotics.

Author Keywords
Human-Robot Interaction; Engagement Awareness; Peripheral Computing Devices; Robot Behaviors.

INTRODUCTION
The use of computing devices, e.g., Personal Computers (PC), mobile phones, wearable devices, in peripheral settings during face-to-face communication has become common in our daily life. These peripheral devices can potentially influence the quality of our conversations [27]. For example, when talking with friends, we may occasionally search for information on our phones, or attend to notifications from social media applications. In these scenarios, our partners may notice such attention shift, and react accordingly, e.g., raising their voice, or pausing until our attention comes back, to prevent possible communication breakdowns.

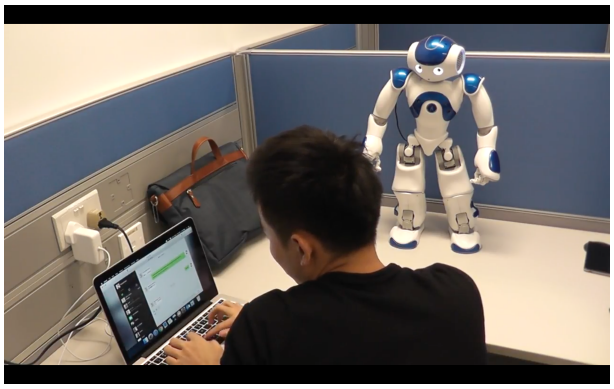


Figure 1: One participant is using the laptop while still maintains interaction with the robot

Similar situations could happen in Human-Robot Interaction (HRI) with the presence of peripheral computing devices, as shown in Figure 1. If the robot cannot tell whether people are still engaged and fails to make proper responses, the interaction flow may be interrupted, leading to potential information loss for human participants.

However, perceiving human partners' engagement dynamics and making proper reactions for robots are challenging. First, when human participants suddenly allocate their attention to peripheral computing devices, the intention is unclear to the robots and the consequent engagement status is ambiguous. For example, when people start to type in laptops, they may be taking relevant notes of the conversation, or replying to irrelevant emails. In the former case, the human participants may still listen to the robots and remain engaged in the conversation, while in the latter, they might be completely out of

CHI 2017, May 06-11, 2017, Denver, CO, USA

 ISBN 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025469

the loop. In other words, different intentions of using computing devices may lead to engagement ambiguities for robots to handle, which is often overlooked in previous research.

Second, for robots, behaving appropriately based on different engagement status of human partners is also a non-trivial task. For instance, the robots can either wait until the humans' attention comes back or actively raise their voices to regain their human partners' attention. Different strategies may result in disparate human perceptions towards their robot partners, which may ultimately have an impact on the quality of the communication. In interpersonal interactions, people have formulated various strategies to make the conversation smoother and more effective [13]. However, whether and how robots can adopt these similar strategies is still under investigated.

In this paper, we consider the computing devices as social beings which can compete for humans' attention with robots. Under this assumption, we define four new types of engagement status based on the human participants' social signals, such as gaze, head pose, and voice. The new types of engagement status offer a fine-grained description of humans' engagement dynamics. We further build a real-time engagement inference model, which captures the inter-state transitions and adjusts the weights on different social signals dynamically according to the involvement of the computing devices. In addition, based on behavioral coding results of a Human-Human Interaction video corpus in a similar configuration with peripheral computing devices present, we design two behavior strategies (*explicit* and *implicit*) for robots to handle the ambiguity problems and ensure successful Human-Robot communication. To evaluate the effectiveness and usefulness of our model and strategy designs, we conduct a within-subject user study of HRI involving a laptop as the peripheral computing device with 27 participants, and analyze their experiences and perceptions through questionnaires and interviews. The participants reflect that the robots with our engagement inference model can perceive their engagement status changes, and are capable of mitigating potential communication breakdowns compared to the one without. In addition, our analysis results reveal that the robots with the *implicit* and *explicit* behavior strategies can handle the information delivery tasks equally well; however, the *implicit* robot is considered to be more polite and appropriate. We also give some implications for behavior designs of engagement-aware robots based on our experimental findings.

RELATED WORK

Peripheral Computing in Interpersonal Communication

The involvement of computing devices in interpersonal communication has been studied for many years. Newman and Smith [28] explored laptop usage in meeting scenarios. They found that laptop users are more likely to drift their attention to less relevant tasks on the laptops, and have difficulty in re-engaging the conversation. In their work, the computing devices, i.e., laptops, are regarded as distractions of the communication. However, the computing devices can also serve as the facilitators for smooth communication. For example, Teevan *et al.* [39] designed a system that allows the effective interaction between presenters and audiences. By building channels to show real-time aggregated feedback of audiences through mobile phones, the presenters are dynamically connected with the audience. And the audiences feel more engaged when they can instantly provide feedback to the presenter. Similarly, Mattias *et al.* [5] studied phone use during meetings and explored its potential influence among attendees. Through experiments in real meetings, they concluded that proper using of smart phones can keep participants' attention on meeting-related tasks and the smart phones can achieve a collaborative presence in meetings. In addition, Hoffman *et al.* [15] showed that peripheral computing devices can also enhance face-to-face interactions as an empathy-evoking supplement. They designed an ambient lamp that can monitor and respond to ongoing conversations on the side. Study showed that participants find the lamp more as a companion than a distraction to the interpersonal communication. Other similar phenomena of using computing devices have also been studied in education and family scenarios [10, 14, 2, 29], where the pros and cons of computing device usage are widely discussed.

Over the years, researchers tried to explore human participants' perceptions and responses to peripheral computing devices in many different application scenarios. Some research on augmented interactive room [17] and ambient displays [42] showed that such systems can act as an additional information channel that may divert users' attention from the focus tasks. Other studies explore the use of ordinary computing devices in peripheral settings. Greatbatch *et al.* [13] considered computing devices in medical consultant scenarios and described how patients coordinate their responses when doctors are using computers. Their research suggests that patients have the impulse to figure out doctor's engagement status, e.g., asking for response, within a certain period. Iqbal *et al.* [16] studied the usage of computing devices, e.g., mobile phones and laptops, in presentations and explored the cost and attitudes about the usage. They reported that audiences worry about missing certain amount of information when using the computing devices. And the presenters are concerned about "whether the device is used in a positive way (e.g., taking notes) or a negative way (e.g., distract by instant messages)". Oduor *et al.* [29] reported people's responses to the usage of mobile devices in home settings. Their results show that people tend to guess what activities the other members are doing when they are using mobile devices. If the activities are relevant to the conversation, e.g., searching locations for family picnics when talking about weekend plans, the usage of these devices is acceptable and beneficial. Otherwise, such usage may lead to frustrations and breakdowns of the conversation. All these studies on computing device usage yields an insight into their influence on participants' engagement status in interpersonal communication. However, little research has explored the effect of computing device usage in HRI. Previous research look mostly into multiparty HRI with a pure face-to-face conversation, in reality, such HRI might involve other computing devices which could complicate the situation. First, the robot in HRI with computing devices situated has no access to the content on these peripheral devices, making it hard to assess their relevance to the on-going conversation. However, the relevance may be inferred from the speeches of human partners in multiparty face-to-face interaction. Second,

the robot may have little clue on the start and end time of the attention shift caused by the computing devices, while, in multiparty interactions, such information can be predicted by leveraging turn-taking signals. Furthermore, how the robot should respond to human partner's use of computing devices to avoid potential communication breakdowns still remains under investigated.

Engagement Measurements

In HRI community, engagement is defined as "the process by which interactors start, maintain, and end their perceived connections to each other during an interaction" [35]. Studies of social and service robots, e.g., museum robots [45], bartender robots [12], dialog agents [6], often assume full engagement status from human participants during interaction and allow few distractions from peripheral computing devices. Under this assumption, previous research focused on different periods of "engagement", i.e., "start", "maintain" and "end", and proposed multiple methods to infer engagement status of human partners. Bohus *et al.* [6, 8] built forecasting models to predict people's intents to start or end the conversation in a situated dialog system. Yamazaki and other researchers [45, 38, 41, 24] focused more on determining people's willingness to maintain the interaction in education and entertainment scenarios.

Most of existing engagement inference models take humans' social signals, e.g., gaze, head pose, body orientation, etc., as input. Among all the social signals, gaze is considered to be one of the most efficient indicators [23]. And gaze related features, like eye contact [44], gaze direction [18], gaze pattern [25], are widely adopted to determine the engagement status of human participants. But since the precise gaze estimation requires expensive equipment, researchers have leveraged the head pose as an alternative cue for engagement measurement [31, 1, 3, 21, 41, 26, 34, 20]. Combining multiple cues to get more accurate estimation of engagement is also adopted in [34, 7, 33, 43, 8, 4, 11], and such combination can provide better engagement estimation than the gaze cue alone [33]. Moreover, some models also incorporate other interaction features, e.g., contextual information [11], conversational history [30], to further improve the engagement inference. However, one potential drawback of the cue-combined methods is that the weights of different cues are often fixed. When human partners in HRI are fully engaged in the conversation, gaze and head pose are the strongest indicators of their engagement status. But if they divert their gaze to the peripheral computing devices, the significance of gaze and head pose for engagement may be downplayed by other cues, e.g., head motion (nodding, etc.), voice feedback, and so on. With the fixed weights on multiple cues, the conventional model can hardly tackle this problem. Furthermore, most of the cue-based measurements of engagement for human partners in HRI are binary [43], (engaged or disengaged), or scaled (based on the engagement strength [24]). Such binary or scaled measurements can hardly describe the dynamic process of engagement. Engagement is expected to "wax and wane" over the interaction [24], and this is especially true when human participants are using computing devices while still involved in the interaction. They might repeatedly divert their attention between the devices

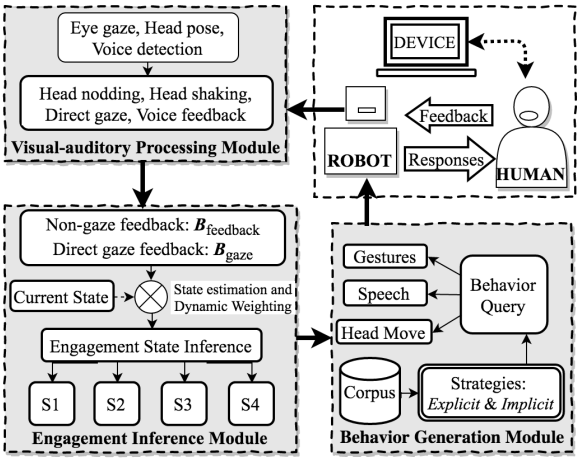


Figure 2: System overview.

and other participants, and dynamically modify their behaviors, e.g., nodding. In such situations, the engagement status of human partners is changing, but it is difficult to infer the corresponding binary values or scaled engagement strengths.

Ambiguity Handling Strategy Design

Ambiguity handling have been studied for decades. For example, in spoken dialog system, the conversational agents may encounter various ambiguities caused by understanding errors. These uncertainties have a negative impact on the dialog performance. If the dialog agents fail to respond appropriately, the communication flows are likely to be disrupted. Researchers in this domain have designed and evaluate multiple strategies for the dialog agents. For example, Bohus and Rudnicky [9] experimented 10 strategies, including repeat asking, further inquiring, uncertainty claiming, etc., for the agents to resolve error-induced ambiguity. Their results show that the further inquiring can achieve the best performance. Marsi and Rooden [22] designed audio-visual expressions, e.g., audio with facial expressions, as well as audio-alone expressions for the embodied agent with a talking head to express confusions. The comparison between the two expressions suggests that the ambiguity can be more reliably resolved via audio-visual expressions. Pejsa *et al.* [32] created two sets of strategies, i.e., speaking policy and listening policy, to stress uncertainty information for a situated conversational agent. The virtual agent can choose policy based on the detected uncertainty state. These studies concentrate on the dialog uncertainties and consider more about the misunderstandings rather than the engagement ambiguity. Yet the proposed strategies can be inspiring for our robot strategy design.

MODEL AND SYSTEM DESIGN

In this section, we introduce our engagement inference model and robot behavior generation in detail.

System Overview

Our system [1] consists of three components: *Visual-auditory Processing*, *Engagement Inference*, and *Behavior Generation*,

[1] Downloadable in https://hcihkust.github.io/EngageDynamics/

as shown in Figure 2. During the interaction with human partners, the robot keeps capturing and analyzing human participants' social signals, e.g., head pose, voice, etc., in the *Visual-auditory Processing* module. The results are then fed into the *Engagement Inference* module, which infers human engagement states in real time by adjusting the weights of the detected social signals dynamically. These states are further used in the *Behavior Generation* module to guide the robot's behaviors according to our predefined behavior codebook[2]. The detailed descriptions of each module are presented in the following subsections.

Visual-auditory Processing

The *Visual-auditory Processing* module exploits the toolkit provided by Aldebaran[3] to capture basic social signals, including gaze direction, head pose and voice information. We then extract high-level semantic social indicators, such as head nodding $B_{headnod}$, head shaking $B_{headshake}$, direct gaze B_{gaze}, and voice B_{voice}, from the low-level social signals as follows:

Head nodding/shaking. We denote the head pose as $\mathbf{v}_h^{(t)} = [yaw^{(t)}, pitch^{(t)}, roll^{(t)}]$, where $yaw^{(t)}$, $pitch^{(t)}$, $roll^{(t)}$ are yaw, pitch, roll angles in the robot coordinate system at time t, respectively. To detect head nodding indicator $B_{headnod}$ at time t, we calculate the absolute deviation of pitch angles within the past N data points: $\delta_{nod} = \frac{1}{N}\sum_{n=0}^{N-1} |pitch^{(t-n)} - \mu_{nod}|$, where $\mu_{nod} = \frac{1}{N}\sum_{n=0}^{N-1} pitch^{(t-n)}$. In our experiment, we set $N = 50$ with sampling rate 25fps (frame per second) according to the pilot study results. If $\delta_{nod} > 0.4$, we set $B_{headnod} = 1$, otherwise $B_{headnod} = 0$. Similarly, we use the yaw angles to obtain the head shaking indicator and set $B_{headshake} = 1$ if the head shaking is detected, otherwise, $B_{headshake} = 0$.

Direct gaze. We use the gaze indicator directly from Aldebaran gaze analysis toolkit, which computes the gaze direction angles $[yaw^{(t)}, pitch^{(t)}]$ relative to the plane of the participant's face. If the participant is looking at the robot, we set $B_{gaze} = 1$, otherwise, $B_{gaze} = 0$.

Voice. We use a 20ms (millisecond) buffer to store the latest raw sound segment. If the total energy of this buffer exceeds a predefined threshold (obtained from the pilot study results), the voice signal is assumed to be detected, $B_{voice} = 1$, otherwise, $B_{voice} = 0$.

Engagement Inference

To reflect humans' engagement dynamics, we define four engagement states as shown in Figure 3 based on two sets of clues. One is direct gaze, B_{gaze}, and the other is the collective non-gaze clues, $B_{feedback} = B_{headnod} \vee B_{headshaking} \vee B_{voice}$, where $\vee$ is the logical *or* operation. The four states, S1, S2, S3 and S4, are defined as follows. The state S1 denotes that the participants are fully engaged as they are looking at the robot, $B_{gaze} = 1$, e.g., **a** in Figure 3. The state S2 denotes that the human participant shifts gaze away, $B_{gaze} = 0$, but is still engaged as indicated by the non-gaze feedback signals, $B_{feedback} = 1$, e.g., **b** in Figure 3. The state S3 denotes that the participants do not show any feedback signals, i.e., $B_{gaze} = 0$ and $B_{feedback} = 0$, for less than 6s (second), e.g., **c** in Figure 3. The state S4 denotes that the participant is assumed to be disengaged, as no social feedback signals are detected, i.e., $B_{gaze} = 0$ and $B_{feedback} = 0$, for more than 6s, e.g., **d** in Figure 3.

[2]Downloadable in https://hcihkust.github.io/EngageDynamics/

[3]http://doc.aldebaran.com/2-1/naoqi/index.html

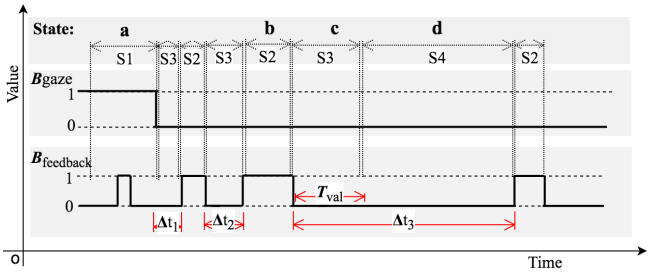
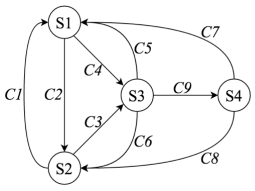


Figure 3: The engagement state S1, S2, S3 and S4. Δt_1, Δt_2 and Δt_3 are time duration of no feedback signals. T_{val} is a predefined threshold (6 seconds) to trigger S3 to transit to S4.

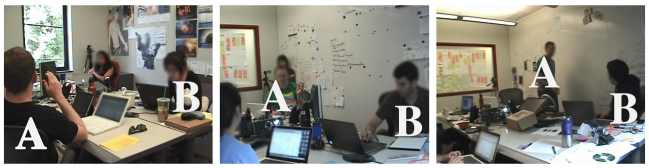


Conditions	Descriptions
C1, C5, C7	$B_{gaze} = 1$
C2, C6, C8	$B_{gaze} = 0, B_{feedback} = 1$
C3, C4	$B_{gaze} = 0, B_{feedback} = 0$
C9	$B_{gaze} = 0, B_{feedback} = 0, \Delta t > 6s$

Figure 4: The transitions of engagement states

We further introduce transitions between these predefined states as illustrated in Figure 4 to explain the underlying principles of the engagement inference. In state S1, the participants' engagement is most certain since they are looking at the robot. From S1 to S2/S3, the gaze has been diverted and the robot has to rely on the non-gaze cues to estimate participants' engagement. Thus, we put more weights on the non-gaze indicator $B_{feedback}$. In other words, the non-gaze cues (head nodding, head shaking, voice) are the only factor to differentiate whether a participant is in S2 or S3. The robot will also assume the participants to be engaged if any non-gaze cues are detected, $B_{feedback} = 1$. Conversely, in state S1, more weights will be put on the gaze indicator B_{gaze}, as the value of $B_{feedback}$ has little impact on this state.

If the participants do not look at the robot and provide no other signals for a short while, i.e., in state S3, there are no clear evidences to determine participants' engagement. Given that the next transition state of S3 could either be engaged, e.g., S1 and S2, or disengaged, e.g., S4, the robot would need additional information to better disambiguate the engagement status in S3. Hence we design a set of handling strategies for the robot, to which the participants' reactions can provide more clues to their actual engagement status, and may ultimately affect the quality of the communication. In the following subsection, we provide detailed descriptions of engagement handling strategy design.



(a) (b) (c)

Figure 5: Snapshots from our video corpus: (a) A raises the hand to regain B's attention. (b) speaker A slows down and waits for B. (c) A turns around to look at B to make sure B is fully engaged.

Cues	Example codes
Gaze	Mutual gaze, gaze sweeping...
Eye Movement	Rapid blinking, looking up...
Head Movement	Nod, shake, toss...
Head Orientation	Turning around, lowering...
Gesture	Pointing, returning to rest position...
Hand Movement	Finger tapping, waving...
Body Orientation	Turning towards, leaning forward...
Body Movement	Stretching, blending...
Posture	Mirroring, standing up...
Speech	Greetings, questions, small talk...
Conjunctions	yeah, OK, well...

Table 1: Cue codes in our codebook

Behavior Generation

The *Behavior Generation* Module is designed to determine what to do and how to do it when different states are detected. Specifically, the robot should exercise state-sensitive behaviors in S1, S2, S3 and S4 to make information delivery smoother and more effective, especially when the engagement status is ambiguous (in S3). To design appropriate behaviors for the robot, we learn from Human-Human Interaction and adapt human strategies in the following three steps.

First, we built a video corpus of multiparty work-related Human-Human Interactions to extract engagement-related actions under different scenes. We shadowed two teams to record their group meetings over a semester. Team A is a course project team with four Master students in Human-Computer Interaction, and Team B is an interest group of three faculties and one Ph.D. student in Design and Human-Computer Interaction. The final corpus, about 850min (minute) in total, contains 17 videos that capture different scenes, such as project planning, presentation rehearsal, design discussion, and meeting arrangement. Some snapshots of the video corpus are shown in Figure 5.

Second, we annotated human behaviors in the video corpus based on a pre-compiled codebook. The codes consist of a variety of essential verbal and non-verbal cues detectable by existing sensors for estimating each conversational participant's engagement status and others' reactions. We construct the initial codebook based on literature review, and further refined it according to sample clips from our video corpus. Table 1 shows examples of cue codes in the codebook. We hired a professional video annotation service to code the participants' engagement status as defined in this paper and their corresponding social signals based on this codebook. The final version of the annotation consists of social signals, the corresponding engagement status, the associated participants, the context, the consequences and the time stamps. Figure 5 shows a few examples: (a) when person A is referring to an important event while his partner B is looking at her laptop, A may infer B is disengaged and thus raise his hand to regain B's attention. (b) The team is discussing their product roadmap when person B suddenly attends to his laptop for a social message, the speaker A therefore slows down and waits. (c) When the speaker A starts an urgent event, he turns around to look at his partner B to imply the importance. These provide essential insights into possible responses the robot should give under the similar circumstances.

Third, by analyzing speakers' behaviors when listeners are possibly disengaged, we find two styles of behavior patterns, perhaps due to different personalities and/or social roles. We denote these two types of strategies as *explicit* and *implicit*, respectively, based on our codebook. The *explicit* strategy expresses views more openly and proactively, such as saying "are your listening?", raising hands to attract attention, etc. The *implicit* strategy will express the speakers' intents in a submissive manner, such as pausing for a moment, saying "I will wait for you", etc. To evaluate the appropriateness and effectiveness of these two strategies for HRI, we conduct a comparative study with a control condition, *unawareness*, which contains no behavior strategies when attention shifts are detected. The detailed behavior manners of the above three strategies are summarized in Table 2.

EXPERIMENT

In order to evaluate the capability of the proposed engagement inference model and different behavior strategy designs, we conduct a within-subject controlled experiment with 27 participants. In this study, we measure the participants' responses and perceptions during dyadic conversations with a robot assistant in work settings where peripheral computing devices may sidetrack humans' attention. More specifically, we use the robot default mode without any behavior strategy as the control condition, denoted as *unawareness*. We design two versions of robot assistants with our engagement inference model but employ different behavior strategies: *explicit* versus *implicit* (see section Model and System Design for more details).

In the experiment, each participant interacts with three versions of robots separately to explore how well the robot can handle engagement dynamics caused by the involvement of peripheral computing devices. To minimize learning and order effects, we counterbalance the order of the three designs of robots as well as their assignments to three different scenarios in a work environment, i.e., *Morning Report*, *Arrange Meetings*, and *Daily Summary*.

Table 3 shows the snippets of the scripts for different scenarios and examples of responses from the three versions of robots

Engagement States	The Robot's Behaviors		
	Explicit	*Implicit*	*Unawareness*
State $S1$: The user is looking at the robot.	The robot talks without pauses	The robot talks without pauses	The robot talks without pauses
State $S2$: The user stops looking at the robot but shows feedback signals.	The robot shifts its attention based on the user's head pose and begin talking slowly.	The robot shifts its attention based on the user's head pose and begin talking slowly.	The robot continues talking as before.
State $S3$: The user attends to the laptop and shows no feedback signals for a short period.	The robot will wait for a while and then directly ask the user's engagement, e.g., "Are you still with me?", "Are you listening?", "Are you following me?", etc. Meanwhile, the robot will also adjust its posture to the partner when asking.	The robot will wait for a while and then pop out filler words, e.g., "Mmm", "Uhm", "Well", "Fine". Meanwhile, the robot will also move its head to follow the user's attention.	The robot continues talking as before.
State $S4$: The user attends to the laptop for quite a long time without any feedback.	The robot will raise the hand to regain attention from the user, e.g., "Hey, listen to me, it's important!", "Hey, you don't want to miss this information!". After that the robot will repeat the last few sentences.	The robot will stop the current conversation and wait for the user to finish the tasks by saying "No problem, I will wait for you.", "OK, I will pause for a moment.", etc.	The robot continues talking as before.

Table 2: The behavior strategies of our three versions of robot assistants under different engagement states.

Scenario	Script Samples
Morning Report	"You have a meeting with Amy at 10am." "You will talk with Amy to decide whether you are going to hike there."
Arrange Meetings	"Davis wants to meet you at twelve o'clock this Tuesday!" "Dr. Wang emailed to confirm next week's meeting arrangement.'
Daily Summary	"Today you have finished your math assignments." "You have to prepare for tomorrow's language course presentation."

Table 3: Script samples in our experiment.

in response to various events. We treat the different robot behavior strategies as our independent variable, and evaluate them in terms of the quality of the service, the perceptions of robots, and user experience.

Hypotheses

Our engagement inference model can detect and disambiguate dynamics in human engagement shifts, based on which robots can respond in different manners. Previous works show that in Human-Human Interaction exercising engagement awareness behaviors can make conversations smoother and more comfortable [32]. Therefore, we hypothesize that:

H1. Robots with our engagement inference model, regardless of their actual styles of behaviors, can better mitigate communication break downs caused by peripheral computing devices. More specifically, human partners will (*H1a*) experience significantly less information loss, (*H1b*) be significantly easier to resume the conversation after using the computing devices, and (*H1c*) feel that the conversation is significantly smoother overall.

H2. Robots with our engagement inference model in general are (*H2a*) perceived to be significantly more competent as an assistant in work settings than those without, and (*H2b*) are significantly more welcomed for future usage.

For manipulation check of our engagement awareness model, we ask people if the robot assistants are aware of their changes of attention on a 7-point Likert scale.

We further hypothesize that different manners of engagement inference robot behaviors can result in different human perceptions of the robots. More specifically:

H3. Compared to those with *explicit* strategy, robots using implicit strategy are perceived to be significantly (*H3a*) less annoying, (*H3b*) less controlling, (*H3c*) less aggressive, (*H3d*) more considerate, (*H3e*) more appropriate, and (*H3f*) more polite.

In our study, we measure these different aspects (derived from [37], [19], [40], [36]) on a 7-point Likert scale.

Experiment Design and Setup

In our experiment, an Aldebaran Nao [4] serves as the robot agent, and we utilize its embedded functions to detect head pose, gaze, and voice. In addition, we use the default implementation as the baseline for the *unawareness* mode. In the other two engagement sensitive modes, the robot captures the social signals from the participant through its visual and audio sensors, and sends the data via WiFi to a back-end laptop (Intel CPU 2.3GHz, 10GB RAM, Ubuntu 14.04) which runs

[4] https://www.ald.softbankrobotics.com/en/cool-robots/nao

our algorithms to estimate the engagement status. Based on the results, the laptop sends command to the robot to adjust its behaviors according to the predefined strategy on the fly.

During the experiment, the participant is asked to sit up straight in a cubic with the Nao robot standing on the table on their right hand side. To the participant's left, a laptop (MacBook Pro with Retina display 13", Intel CPU 2.7GHz, 8GB RAM) serves as the peripheral computing device. As shown in Figure 1. The Nao robot is roughly the same height as the participant in sitting position, and the distance between the robot and its partner is between 0.4m and 1.5m for better social signal detection. Note that the distance may vary as the participant may change the body orientation and/or posture. The Nao robot can rotate, tilt and bend its head and upper body actively, and can also gesture with its hands. However, we do not allow the robot to walk around on the table to avoid unintentional interference with the conversations or tasks due to sudden movements.

The study consists of three sessions, each involving a different version of the robot (*explicit*, *implicit*, and *unawareness*) in one of the three different scenarios (*Morning Report*, *Arrange Meetings*, and *Daily Summary*). In each session, while the participants are interacting with the robot in English, an experimenter sitting in another cubic sends two requests for the participants to complete from the back-end computer to the laptop on the participant side via a social messaging application at random. On the participants' machine, the requests automatically pop up with a notification sound. The requests can be relevant (e.g., "describe the professor as much as possible") or irrelevant (e.g., "what's your favorite app in your mobile phone?') to the conversation at the time.

With participants' consent, we record the whole experiment by a Panasonic video camera.

Participants

We recruit 27 volunteers from a local university (seven females, average age: 24, SD: 1.72) by word-of-mouth and flyers. According to the pre-screening survey, 10 of them report that they have some experiences of interacting with physical or virtual conversational robots, such as Apple Siri [5]. All participants are university students with different education background. They all have TOEFL score above 95 or IELTS score above 7.0, which indicates that they have no problem communicating with the robot in English.

Procedure

After obtaining the consents from participants, we introduce the procedure of the experiment. Each participant takes part in three sessions in a counterbalanced order. In each session, the robot first introduces itself, has some small talks with the participants, and then start the task-related topics. During the conversations, the participant attends to the laptop on the side twice upon the arrival of requests through the messaging application. Each session lasts for less than 10 minutes depending on individual speed of completing the requests. At the end of each session, we ask the participant to fill out a questionnaire

[5] http://www.apple.com/ios/siri/

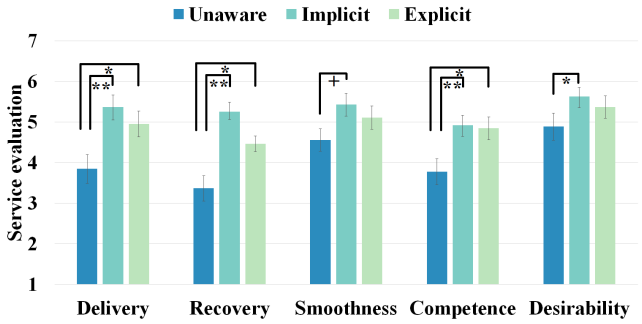


Figure 6: Means and standard errors of the service evaluation of our robot assistants on a 7-point Likert scale (+: $.05 < p < .1$, *: $p < .05$, **: $p < .01$)

to rate the service performance of the robot assistant. Upon the completion of the three sessions, the participant compares and rates the three versions of robots in terms of perceptions and user experience on a 7-point Likert scale, and the participant can review the video recordings if needed. In the end, we conduct an in-depth interview with the participant to find out more about the feelings regarding the robot.

During the pilot study, we detected some transition errors in our model, mostly caused by face/gaze tracking failures. To mitigate this issue, in the experiment, we program the robot to remind the participant to sit up straight or move closer to it, if failed to detect any face or estimate the head pose. In the meantime, we keep monitoring our system log and mark any misbehavior of the robot without interrupting the ongoing experiment. In the post-study interview, we ask additional questions about the user's perceptions and reactions to these incidents to help improve the performance of our system.

ANALYSIS AND RESULTS

We summarize the statistical analyses and interview results, in terms of participants' perceptions of the service and the robot.

Manipulation Check

The manipulation check for engagement awareness conditions shows that the manipulation is effective (repeated measures MANOVA, $F(2, 52) = 37.02$, $p < 0.01$, $\eta^2 = .59$). The robots with our engagement model are indeed perceived as being able to detect participants' engagement dynamics (*explicit*: $M = 5.41$, $SD = 1.65$; *implicit*: $M = 5.67$, $SD = 1.41$) than the one without ($M = 2.67$, $SD = 1.52$); Bonferroni post-hoc test $p < 0.05$.

Service Evaluation

Figure 6 shows the robot's ability to avoid communication break downs due to the presence of peripheral computing devices.

The engagement-aware robots are significantly more capable of minimizing information loss when interruptions occur during their conversations with the human partners than the insensitive one (repeated measures MANOVA, $F(2, 52) = 8.51$, $p < 0.01$, $\eta^2 = .25$; ***H1a*** accepted). The participants feel that both the *explicit* robot ($M = 4.96, SD = 1.65$) and

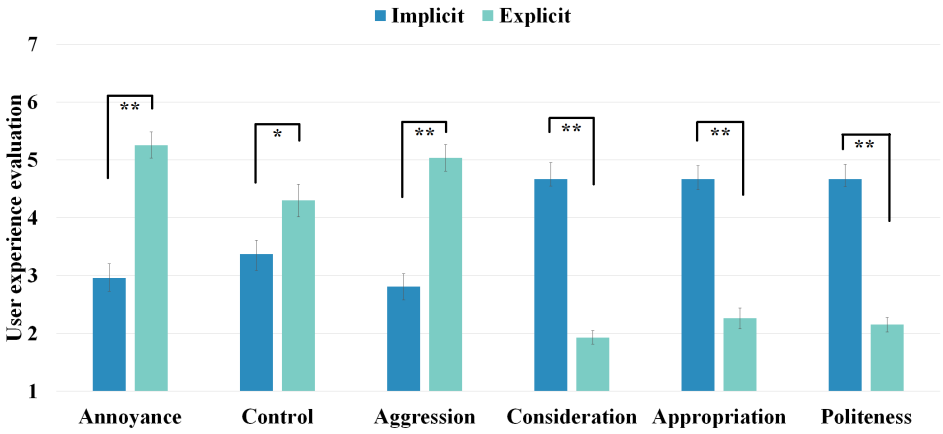


Figure 7: Means and standard errors of the user experience evaluation of our robot assistants on a 7-point Likert scale (+: $.05 < p < .1$, *: $p < .05$, **: $p < .01$)

the *implicit* robot ($M = 5.37, SD = 1.55$) deliver significantly more information than the default *unawareness* one ($M = 3.85, SD = 1.83$); Bonferroni post-hoc test $p < 0.05$. There are no significant differences between the two behavior strategies.

Similarly, participants suggest that they can recover from the interruptions and resume the conversation significantly more easily with the robots running our engagement inference model (repeated measures MANOVA, $F(2, 52) = 13.75$, $p < 0.01$, $\eta^2 = .35$; ***H1b*** accepted). Both the *explicit* ($M = 4.74, SD = 1.48$) and the *implicit* ($M = 5.26$, $SD = 1.43$) robots can resolve uncoupling significantly better than the baseline robot ($M = 3.37$, $SD = 1.45$); Bonferroni post-hoc test $p < 0.05$.

Furthermore, the conversations with engagement-aware robots are significantly smoother than that with the ordinary robot ($F(2, 52) = 3.61$, $p < 0.05$, $\eta^2 = .12$; ***H1c*** accepted). Bonferroni post-hoc check shows that handling attention shift using *implicit* strategy ($M = 5.44$, $SD = 1.22$) can lead to marginally smoother conversations than the baseline robot ($M = 4.56$, $SD = 1.63$; $p = 0.07$). But the effect of the explicit strategy ($M = 5.11$, $SD = 1.01$) is not significant.

In terms of competence and desirability, results show that the participants view the robots with our engagement inference model as more competent than the ones without. (repeated measures MANOVA, $F(2, 52) = 8.01$, $p < 0.01$, $\eta^2 = .24$; ***H2a*** accepted) Bonferroni post-hoc test suggests that while behavior strategies have no significant impact on the perceived competence, participants feel that the robots with our engagement model (*implicit*: $M = 4.93$, $SD = 1.21$; *explicit*: $M = 4.85$, $SD = 1.43$) are significantly better at their jobs than the ones without ($M = 3.78$, $SD = 1.72$; $p < 0.05$).

Similarly, participants prefer working with the engagement-aware robots in the future, significantly more than the baseline one (repeated measures MANOVA, $F(2, 52) = 3.25$, $p < 0.05$, $\eta^2 = .11$). Bonferroni post-hoc test further reveals that people prefer significantly more the implicit robot ($M = 5.63$, $SD = 1.28$) than the baseline robot ($M = 4.89$, $SD = 1.63$; $p < 0.05$). But the difference between the explicit robot ($M = 5.37$, $SD = 1.45$) and the baseline one is not significant. Therefore, ***H2b*** is only partially accepted.

Robot Evaluation and User Experience

To further explore the underlying rationales of participants' preferences, we ask them to compare the *explicit* and *implicit* robots in terms of *annoyance*, *controlling*, *aggression*, *consideration*, *appropriateness*, and *politeness* on a 7-point Likert scale. The statistical results are shown in Figure 7.

In general, participants find the explicit robot to be ($M = 5.26$, $SD = 1.20$) significantly more annoying than the implicit one ($M = 2.96$, $SD = 1.26$); repeated measures MANOVA, $F(2, 52) = 17.76$, $p < 0.01$, $\eta^2 = .41$; Bonferroni post-hoc test $p < 0.01$ (***H3a*** accepted). This is also confirmed in the post-study interviews:

> *"I like the first one (implicit) the most. It is more patient. The last one (explicit) is too annoying as it kept talking when I was responding to the requests. It is quite noisy. ... I have no special feeling for the second one (unawareness). ... "* – P12 (Male, age: 26)

Participants feel that the explicit robot ($M = 4.30$, $SD = 1.46$) is significantly more controlling than the implicit one ($M = 3.37$, $SD = 1.25$); repeated measures MANOVA, $F(2, 52) = 3.45$, $p < 0.05$, $\eta^2 = .12$; Bonferroni post-hoc test $p < 0.05$ (***H3b*** accepted). Some participants commented in the interview that:

> *"It (explicit) always said "Are you listening to me?" It persisted until I finally turned back to it. The other robot (implicit) just waited for me."* – P9 (Male, age: 24)

Similarly, the explicit robot ($M = 5.04$, $SD = 1.19$) is more aggressive than the implicit one ($M = 2.81$, $SD = 1.18$); repeated measures MANOVA, $F(2, 52) = 16.40$, $p < 0.01$, $\eta^2 = .39$; Bonferroni post-hoc test $p < 0.05$ (***H3c*** accepted).

Overall, resolving ambiguity and handling disengagement in the implicit manner ($M = 4.67$, $SD = 1.47$) are perceived to be significantly more considerate than that in the explicit manner ($M = 1.93$, $SD = 0.62$); repeated measures MANOVA,

$F(2, 52) = 78.37$, $p < 0.01$, $\eta^2 = .75$; Bonferroni post-hoc test $p < 0.01$ (***H3e*** accepted). One participant mentioned in the interview that:

> *"I really appreciate the robot (implicit) giving me time and space to work on the computer, very thoughtful."* – P20 (Female, age: 23)

Participants thinks the implicit robot ($M = 4.67$, $SD = 1.21$) to be significant more socially appropriate than the explicit one ($M = 2.26$, $SD = 0.94$); repeated measures ANOVA, $F(2, 52) = 56.85$, $p < 0.01$, $\eta^2 = .69$; Bonferroni post-hoc test $p < 0.01$ (***H3e*** accepted). Similar effects are found in the politeness rating. Participants think the the implicit robot ($M = 4.67$, $SD = 1.30$) to be more polite than the explicit one ($M = 2.15$, $SD = 0.66$); repeated measures MANOVA, $F(2, 52) = 58.97$, $p < 0.01$, $\eta^2 = .69$; Bonferroni post-hoc test $p < 0.01$ (***H3f*** accepted).

In the interviews, some participants also mentioned the following points,

> *"I like the last robot (implicit) manner of speaking. It sounds more polite than others, and it handled the interruption quite tactfully ..."* – P21 (Male, age: 24)
>
> *"The last robot (unawareness) simply rushes the conversations. The second robot (implicit) instead takes its time and waits for me, which is respectful. ... "* – P19 (Male, age: 23)

In summary, our engagement inference model is effective in HRI with peripheral computing devices situated. Our model can detect the engagement state transitions of human partners, and adjust the weights of different cues (i.e., head pose, gaze, etc.) to resolve potential ambiguities. This can help the robots conduct smoother and more effective conversations. In addition, although the two versions of our engagement-aware robots of different handling strategies achieve similar service performance and competence ratings, participants have a more positive experience with the *implicit* design. They report the *implicit* robot more considerate and polite, which shows that robot behavior design is critical to successful HRI.

DISCUSSION

In this section, we discuss some insights derived from our study and the limitations of this work.

Implications for Design

Based on data analysis and interview results, we summarize some implications for designing engagement-aware robot behaviors.

Pausing to Acknowledge Computing Devices' Involvement

When robot assistants detect their human partners' attention shift to other devices, regardless of the actual purposes, it is better for the robots to pause for a short moment to acknowledge the event. Then the robots can take subsequent actions according to certain behavior strategies. This gives the participants the sense that the robots are aware of the attention shift, which may result in more effective communication. In addition, most of the participants report that the implicit robots are more "considerate" and "polite" and thus prefer the implicit strategy for HRI.

Robot Behaviors Should Be Context-aware

After a brief pause when the attention shift of human partners is detected, whether robots should continue waiting or should need to urge the human partners to re-enter the conversation could depend on various contextual factors of the HRI, including task-related factors e.g., relevance and urgency, user-related factors e.g., cognitive capacity and emotion, and environment-related factors e.g., existence of other distractions. For example, when the human and robot team is working on an urgent task, it is necessary for the robot to attract its human partner proactively to quickly recover the conversation from interruption. If the task is not as time critical, it might be better for the robot to wait, which could result in smoother and more comfortable communication.

Robot Behaviors Should Be Intent Sensitive

Robots should be sensitive of their human partners' intentions, which might be inferred through the interaction. For example, when robots keep requesting their human partners' attention back to the current interaction while the partners persistently ignore the requests, robots should respect the partners' intents and adjust their behavior strategies.

Consider Potential Negative Experience Caused by Robots

Some participants prefer the unawareness strategy even though they cannot follow up with the conversation. One possible reason may be the "guilty feeling" of the participants. When robots say "are you listening to me?" (*explicit*) or "you are doing other tasks. I will wait for you." (*implicit*), these speeches may make partners feel guilty as if they had caused some troubles. Therefore, we suggest that the potential negative emotional experiences caused by robots' behaviors should be considered when designing engagement-aware robots. In addition, we identify two other types of behaviors that might lead to negative feelings from our participants' feedback.

Repetitive behaviors. Some participants mention that the explicit strategy is quite annoying as the robot keeps saying similar speeches like "are you listening to me?". Although we designed several different statements to regain participants' attention and alternate them randomly during conversation, participants are still likely to experience the repetitive requests if they do not show responses to the robot for a long time. Therefore, having a larger and more expressive vocabulary is needed to avoid the negative experience.

Speaking in constant pace. Some participants say that the *unawareness* robot speaks too fast, and they find it hard to follow the conversation. In fact, all the three versions of robots have same talking speed. However, the explicit and implicit strategies may change the paces based on the interactions, which may make the conversation more comprehensible and engaging.

Limitations

Our work has several limitations. First, our experiment setup contains only one laptop as the peripheral computing device.

However, several participants did use their phones unexpectedly during the experiments. Although our inference model successfully detects these phone use events, we need more systematic study on multi-device situations. Second, our current engagement inference model cannot handle the face occlusion problem. When participants rest their chins on hands, it is difficult for the face detector embedded in the Nao robot to locate their faces and estimate the head poses accurately, hence it is hard to infer the engagement status. Similarly, the face detector in the Nao robot cannot handle large face rotation angles, which can result in unsatisfactory estimation accuracy of our engagement inference model. Third, we do not consider different participants' personalities, and only use fixed robot behaviors for different strategies. If robot assistants can recognize human personality and respond accordingly, smoother and more effective information communication can be expected. Fourth, we do not differentiate active engagement from passive engagement. In this study, there is no specific purpose for the interaction. Hence, the participants listen passively to the robot during interaction for most of the time. However, they may turn into active listeners if asked to complete a quiz based on what the robot has said, and thus are likely to behave rather differently. We will further investigate the effects of these two modes in follow-up studies. Fifth, we restrict the robot's mobility to avoid unintended interference. However, if given more freedom, the robot could try out more expressive engagement handling strategies by incorporating speech, gesture, body languages, and movement in space. Last, in our experiment, the conversational turns between the robot and humans are not balanced, and the dialog system we employ can only handle small talks. The robot is still far from satisfactory to carry out fluent conversations like a real human.

CONCLUSION AND FUTURE WORK

We propose a new real-time engagement inference model by distinguishing engagement level in a find-grained scale. Based on the proposed model, we investigate two different robot behavior strategies. Our experiment results show that our engagement inference model and behavior strategies are useful and effective during Human-Robot Interaction when peripheral computing devices are considered. We also find it meaningful to make robots wait for the participants when their attention is not on the current conversation. Future works could include improving the performance of our engagement inference model in more scenarios and deploying our system in the real HRI.

ACKNOWLEDGMENTS

We thank the WeChat-HKUST Joint Laboratory on Artificial Intelligence Technology (WHAT LAB) grant#1516144-0, and NSF CIFellows grant#1019343, for sponsoring this research.

REFERENCES

1. Jeremy N. Bailenson, Andrew C. Beall, Jack Loomis, Jim Blascovich, and Matthew Turk. 2005. Transformed social interaction, augmented gaze, and social influence in immersive virtual environments. *Human Communication Research* 31, 4 (2005), 511–537. DOI: http://dx.doi.org/10.1093/hcr/31.4.511
2. Louise Barkhuus. 2005. Bring your own laptop unless you want to follow the lecture: Alternative communication in the classroom. *Proceedings of the 2005 international ACM ...* (2005), 140–143. DOI: http://dx.doi.org/10.1145/1099203.1099230
3. Maren Bennewitz, Felix Faber, Dominik Joho, Michael Schreiber, and Sven Behnke. 2005a. Integrating vision and speech for conversations with multiple persons. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS* (2005), 1295–1300. DOI: http://dx.doi.org/10.1109/IROS.2005.1545158
4. Maren Bennewitz, Felix Faber, Dominik Joho, Michael Schreiber, and Sven Behnke. 2005b. Towards a humanoid museum guide robot that interacts with multiple persons. *Proceedings of 2005 5th IEEE-RAS International Conference on Humanoid Robots* 2005 (2005), 418–423. DOI:http://dx.doi.org/10.1109/ICHR.2005.1573603
5. Matthias Böhmer, T. Scott Saponas, and Jaime Teevan. 2013. Smartphone use does not have to be rude. *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services - MobileHCI '13* (2013), 342. DOI: http://dx.doi.org/10.1145/2493190.2493237
6. Dan Bohus and Eric Horvitz. 2009. Dialog in the Open World : Platform and Applications. *Behaviour* (2009), 31–38. DOI:http://dx.doi.org/10.1145/1647314.1647323
7. Dan Bohus and Eric Horvitz. 2010. Facilitating multiparty dialog with gaze, gesture, and speech. *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction on - ICMI-MLMI '10* (2010), 1. DOI: http://dx.doi.org/10.1145/1891903.1891910
8. Dan Bohus and Eric Horvitz. 2014. Managing Human-Robot Engagement with Forecasts and... um... Hesitations. *Proceedings of the 16th International Conference on Multimodal Interaction* (2014), 2–9. DOI: http://dx.doi.org/10.1145/2663204.2663241
9. Dan Bohus and AIexander Rudnicky. 2005. Sorry, I Didn't Catch That! - An Investigation of Non-understanding Errors and Recovery Strategies. *6th SIGdial Workshop on Discourse and Dialogue* (2005). http://www.isca-speech.org/archive
10. Andrea Beth Campbell and Roy P Pargas. 2003. Laptops in the classroom. *ACM SIGCSE Bulletin* 35, 1 (2003), 98. DOI:http://dx.doi.org/10.1145/792548.611942
11. Ginevra Castellano, Andre Pereira, Iolanda Leite, Ana Paiva, and Peter W Mcowan. 2009. Detecting User Engagement with a Robot Companion Using Task and Social Interaction-based Features Interaction scenario. *Proceedings of the 2009 international conference on Multimodal interfaces* (2009), 119–125. DOI: http://dx.doi.org/10.1145/1647314.1647336
12. Mary Ellen Foster, Andre Gaschler, and Manuel Giuliani. 2013. How Can I Help You? Comparing Engagement

Classification Strategies for a Robot Bartender. In *Proceedings of the 15th ACM on International conference on multimodal interaction - ICMI '13*. 255–262. DOI: http://dx.doi.org/10.1145/2522848.2522879

13. David Greatbatch, Paul Luff, Christian Heath, and Peter Campion. 1993. Interpersonal communication and human-computer interaction: an examination of the use of computers in medical consultations. *Interacting with computers* 5, 2 (1993), 193–216.

14. Helene Hembrooke and Geri Gay. 2003. The laptop and the lecture: The effects of multitasking in learning environments. *Journal of Computing in Higher Education* 15, 1 (2003), 46–64. DOI: http://dx.doi.org/10.1007/BF02940852

15. Guy Hoffman, Oren Zuckerman, Gilad Hirschberger, Michal Luria, and Tal Shani-sherman. 2015. Design and Evaluation of a Peripheral Robotic Conversation Companion. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (2015), 3–10. DOI: http://dx.doi.org/10.1145/2696454.2696495

16. Shamsi T. Iqbal, Jonathan Grudin, and Eric Horvitz. 2011. Peripheral computing during presentations. *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11* (2011), 891. DOI: http://dx.doi.org/10.1145/1978942.1979073

17. Hiroshi Ishii, Craig Wisneski, Scott Brave, Andrew Dahley, Matt Gorbet, Brygg Ullmer, and Paul Yarin. 1998. ambientROOM: Integrating Ambient Media with Architectural Space. In *Proceedings of the Conference on Human Factors in Computing Systems: Making the Impossible Possible (CHI'98)*. 173–174. DOI: http://dx.doi.org/10.1145/286498.286652

18. R Ishii, Yi Nakano, and Toyoaki Nishida. 2013. Gaze awareness in conversational agents: Estimating a user's conversational engagement from eye gaze. *ACM Transactions on Interactive Intelligent Systems* 3, 2 (2013), 1–11. DOI: http://dx.doi.org/10.1145/2499474.2499480

19. Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. 2010. Gracefully mitigating breakdowns in robotic services. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2010), 203–210. DOI: http://dx.doi.org/10.1109/HRI.2010.5453195

20. Severin Lemaignan, Fernando Garcia, Alexis Jacq, and Pierre Dillenbourg. 2016. From real-time attention assessment to "with-me-ness" in human-robot interaction. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2016), 157–164. DOI: http://dx.doi.org/10.1109/HRI.2016.7451747

21. Liyuan Li, Xinguo Yu, Jun Li, Gang Wang, Ji-Yu Shi, Yeow-Kee Tan, and Haizhou Li. 2012. Vision-based attention estimation and selection for social robot to perform natural interaction in the open world. *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on* 3 (2012), 183–184. DOI: http://dx.doi.org/10.1145/2157689.2157746

22. Erwin Marsi and Ferdi Van Rooden. 2007. Expressing Uncertainty with a Talking Head in a Multimodal Question-Answering System. *Communication And Cognition* (2007). DOI: http://dx.doi.org/citeulike-article-id:2429269

23. AJung Moon, Daniel M. Troniak, Brian Gleeson, Matthew K. X. J. Pan, Minhua Zeng, Benjamin A. Blumer, Karon MacLean, and Elizabeth A. Croft. 2014. Meet me where I'm gazing: how shared attention gaze affects human-robot handover timing. *Human-Robot Interaction* (2014), 334–341. DOI: http://dx.doi.org/10.1145/2559636.2559656

24. Samer Al Moubayed and Jill Fain Lehman. 2015. Toward Better Understanding of Engagement in Multiparty Spoken Interaction with Children. (2015), 211–218. DOI: http://dx.doi.org/10.1145/2818346.2820733

25. Yukiko I. Nakano and Ryo Ishii. 2010. Estimating user's engagement from eye-gaze behaviors in human-agent conversations. *Proceedings of the 15th international conference on Intelligent user interfaces - IUI '10* (2010), 139. DOI:http://dx.doi.org/10.1145/1719970.1719990

26. Yukiko I. Nakano, Takashi Yoshino, Misato Yatsushiro, and Yutaka Takase. 2015. Generating Robot Gaze on the Basis of Participation Roles and Dominance Estimation in Multiparty Interaction. *ACM Transactions on Interactive Intelligent Systems* 5, 4 (2015), 1–23. DOI: http://dx.doi.org/10.1145/2743028

27. William Newman. 2006. Must electronic gadgets disrupt our face-to-face conversations? *Interactions* 13, 6 (2006), 18. DOI:http://dx.doi.org/10.1145/1167948.1167968

28. William Newman and EL Smith. 2006. Disruption of Meetings by Laptop Use: Is There a 10-Second Solution? *CHI'06 Extended Abstracts on Human Factors in . . .* (2006), 1145–1150. DOI: http://dx.doi.org/10.1145/1125451.1125667

29. Erick Oduor, Carman Neustaedter, William Odom, Anthony Tang, Niala Moallem, Melanie Tory, and Pourang Irani. 2016. The frustrations and benefits of mobile device usage in the home when co-present with family members. (2016). DOI: http://dx.doi.org/10.1145/2901790.2901809

30. Catharine Oertel, Kenneth A. Funes Mora, Joakim Gustafson, and Jean-Marc Odobez. 2015. Deciphering the Silent Participant. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15*. 107–114. DOI: http://dx.doi.org/10.1145/2818346.2820759

31. Kazuhiro Otsuka, Yoshinao Takemae, and Junji Yamato. 2005. A probabilistic inference of multiparty-conversation structure based on Markov-switching models of gaze patterns, head

directions, and utterances. *Proceedings of the 7th International Conference on Multimodal Interfaces* (2005), 191–198. DOI: http://dx.doi.org/10.1145/1088463.1088497

32. Tomislav Pejsa, Dan Bohus, Michael F. Cohen, Chit W. Saw, James Mahoney, and Eric Horvitz. 2014. Natural Communication about Uncertainties in Situated Interaction. *International Conference on Multimodal Interaction* (2014), 283–290. DOI: http://dx.doi.org/10.1145/2663204.2663249
33. Julia Schwarz, Charles Claudius Marais, Tommer Leyvand, Scott E. Hudson, and Jennifer Mankoff. 2014. Combining body pose, gaze, and gesture to determine intention to interact in vision-based interfaces. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (2014), 3443–3452. DOI: http://dx.doi.org/10.1145/2556288.2556989
34. Samira Sheikhi and Jean Marc Odobez. 2015. Combining dynamic head pose-gaze mapping with the robot conversational state for attention recognition in human-robot interactions. *Pattern Recognition Letters* 66 (2015), 81–90. DOI: http://dx.doi.org/10.1016/j.patrec.2014.10.002
35. Candace L. Sidner, Christopher Lee, Cory D. Kidd, Neal Lesh, and Charles Rich. 2005. Explorations in engagement for humans and robots. *Artificial Intelligence* 166, 1-2 (2005), 140–164. DOI: http://dx.doi.org/10.1016/j.artint.2005.03.005
36. Vasant Srinivasan and Leila Takayama. 2016. Help Me Please: Robot Politeness Strategies for Soliciting Help From Humans. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), 4945–4955. DOI: http://dx.doi.org/10.1145/2858036.2858217
37. Megan Strait, Lara Vujovic, Victoria Floerke, Matthias Scheutz, and Heather Urry. 2015. Too Much Humanness for Human-Robot Interaction: Exposure to Highly Humanlike Robots Elicits Aversive Responding in Observers. *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems* 1 (2015), 3593–3602. DOI: http://dx.doi.org/10.1145/2702123.2702415
38. Daniel Szafir and Bilge Mutlu. 2012. Pay attention! Designing Adaptive Agents that Monitor and Improve User Engagement. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (2012), 11. DOI: http://dx.doi.org/10.1145/2207676.2207679
39. Jaime Teevan, Daniel Liebling, Ann Paradiso, Carlos Garcia Jurado Suarez, Curtis von Veh, and Darren Gehring. 2012. Displaying mobile feedback during a presentation. *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services - MobileHCI '12* (2012), 379. DOI: http://dx.doi.org/10.1145/2371574.2371633
40. Cristen Torrey, Susan R. Fussell, and Sara Kiesler. 2013. How a robot should give advice. *ACM/IEEE International Conference on Human-Robot Interaction* (2013), 275–282. DOI: http://dx.doi.org/10.1109/HRI.2013.6483599
41. Marynel Vázquez, Aaron Steinfeld, Scott E Hudson, and Jodi Forlizzi. 2014. Spatial and Other Social Engagement Cues in a Child-robot Interaction: Effects of a Sidekick. *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction* (2014), 391–398. DOI:http://dx.doi.org/10.1145/2559636.2559684
42. Daniel Vogel and Ravin Balakrishnan. 2004. Interactive public ambient displays. *Proceedings of the 2004 ACM Symposium on User Interface Software and Technology* (2004), 137. DOI: http://dx.doi.org/10.1145/1029632.1029656
43. Qianli Xu, Liyuan Li, and Gang Wang. 2013. Designing engagement-aware agents for multiparty conversations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (2013), 2233–2242. DOI: http://dx.doi.org/10.1145/2470654.2481308
44. Tian (Linger) Xu, Hui Zhang, and Chen Yu. 2016. See You See Me: The Role of Eye Contact in Multimodal Human-Robot Interaction. *ACM Transactions on Interactive Intelligent Systems* 6, 1 (2016), 1–22. DOI: http://dx.doi.org/10.1145/2882970
45. Keiichi Yamazaki, Akiko Yamazaki, Mai Okada, Yoshinori Kuno, Yoshinori Kobayashi, Yosuke Hoshi, Karola Pitsch, Paul Luff, Dirk Lehn, and Christian Heath. 2009. Revealing Gauguin:Engaging Visitors in Robot Guide's Explanation in an Art Museum. (2009). DOI: http://dx.doi.org/10.1145/1518701.1518919