Out of my way! Exploring Different Modalities for Robots to Ask People to Move Out of the Way

Natalie Friedman[1], David Goedicke[2], Vincent Zhang[3], Dmitriy Rivkin[4],
Michael Jenkin[5], Ziedune Degutyte[6], Arlene Astell[7] Xue Liu[8], Gregory Dudek[9]

***Abstract*— To navigate politely through social spaces, a mobile robot needs to communicate successfully with human bystanders. What is the best way for a robot to attract attention in a socially acceptable manner to communicate its intent to others in a shared space? Through a series of in-the-wild experiments, we measured the social appropriateness and effectiveness of different modalities for robots to communicate to people their intended movement, using combinations of visual text, audio and haptic cues. Using multiple modalities to draw attention and declare intent helps robots to communicate acceptably and effectively. We recommend that in social settings, robots should use multiple modalities to ask people to get out of the way. Additionally, we observe that blowing air at people is a particularly suitable way of attracting attention.**

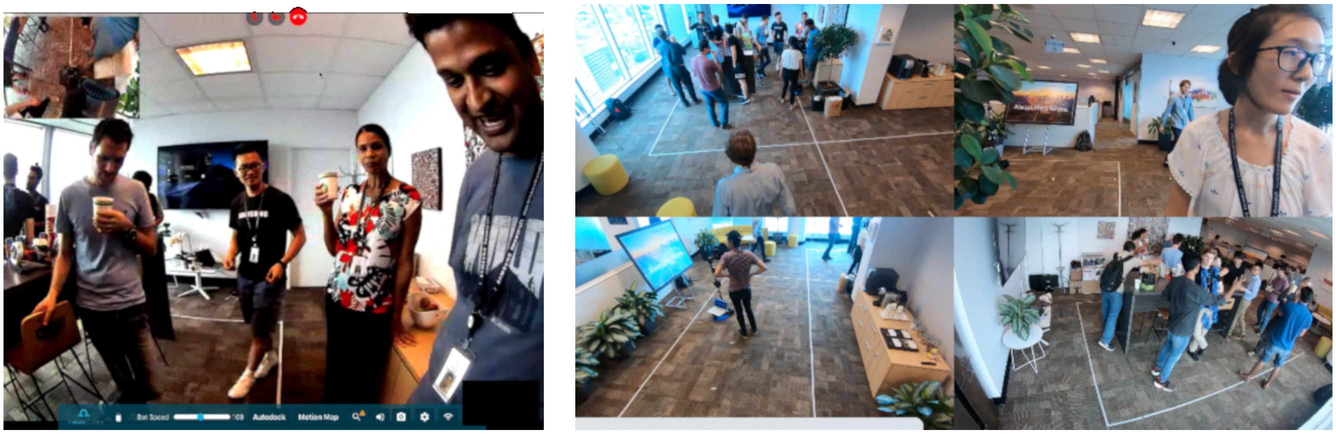
I. Introduction

Humans rely on a wide range of interaction strategies, including motion, touch, semantic free utterances[46], voice, and gaze, to solicit attention when moving through a crowded space [22]. People provide information about their intent through a wide range of cues including body position, walking style, and even by lightly tapping someone on their shoulder and saying "Excuse, me!" Similarly, robots wishing to navigate through densely populated environments must find strategies to gain attention and communicate intent. When the path forward is blocked, the robot must interact with people nearby to capture attention and communicate the desire that people move out of the way. This process should be accomplished in a socially understandable manner. What would be an effective mechanism to communicate this intent in a socially acceptable manner? We adopted two fundamental principles to follow for a robot seeking assistance in moving through a crowd. The robot should: (i) get the right amount of attention at the right time from the people obstructing its path, and (ii) compel people to move out of the way once the robot has their attention. We conducted in-the-wild experiments [35] at informal social gatherings (i.e. office parties) to evaluate attention-seeking behaviors for a mobile robot. These experiments focus on the all too common scenario where the robot's path is blocked by a person.

Although there is certainly a place for controlled laboratory experiments in the study of human-robot interaction, there exist questions that lab studies cannot answer but that in-the-wild studies can naturally answer. As described in [24] in-the-wild experiments provide an opportunity to more fully understand how humans interact with robots in less structured and more natural interaction spaces. They also provide an opportunity to explore unintended consequences of human-robot interactions. While uniquely informative, the organic nature of in-the-wild studies are not without their compromises and concerns. In natural interactions, it is difficult to ensure that the group sizes are identical or that subject pools are completely independent across conditions. This limits the use of strong statistical tests and typically restricts studies to more observational results. Nevertheless, in-the-wild experiments provide useful insights into the design of human-robot interactions and the technologies and methodologies that support them.

We used an "office party" social setting to conduct an in-the-wild study in which a mobile robot was teleoperated to move through a crowd. These events were associated with robotics seminars where most subjects had previous exposure to robots, such that novelty effects and acclimatization overhead were minimized. We used a Wizard of Oz (WoZ) methodology whereby participants had the impression that the robot was moving autonomously following a line on the floor, while it was actually being controlled remotely [26], [9]. This allowed the robot to safely avoid colliding with people while prototyping the system behavior before building a fully autonomous system. We equipped the tele-operated robot with different modalities to attract the attention of people who were standing in its path (a white line on the floor) and to indicate for them to move out of its way (Figure 1). The motion of the robot and the signals used to interact with the participants were remotely triggered by a researcher. Responses to three combinations of different attention-seeking modalities (haptic, visual, audio) were collected during the social events. The haptic cue was a directed wind event created by a fan mounted on the robot, the visual text cue was presented via text on

[1]Natalie Friedman is with the Samsung AI Center, Montreal, Canada nvf4@cornell.edu
[2]David Goedicke is with the Samsung AI Center, Montreal, Canada dg536@cornell.edu
[3]Vincent Zhang is with the Samsung AI Center, Montreal, Canada vincentzhang2244@gmail.com
[4]Dmitriy Rivkin is with the Samsung AI Center, Montreal, Canada d.rivkin@samsung.com
[5]Michael Jenkin is with York University, Canada, and the Montreal Samsung AI Center, Canada m.jenkin@samsung.com
[6]Ziedune Degutyte is with the Cambridge Samsung AI Center, UK z.degutyte@samsung.com
[7]Arlene Astell is with the Cambridge Samsung AI Center, UK a.astell@samsung.com>
[8]Xue Liu is with the Samsung AI Center, Montreal, Canada steve.liu@samsung.com
[9]Gregory Dudek is with the Samsung AI Center, Montreal, Canada greg.dudek@samsung.com



(a) (b)

Fig. 1. (a) the view from the robot's camera on the operator's control display. (b) a quad view of the party space including the robot's planned path marked by white tape.

a tablet screen mounted on the robot, and the audio cue was a noise emitted from the robot. Behavioral responses from participants in the experiment, and post-interaction questionnaire responses were used to gauge the effectiveness and social appropriateness of the different strategies.

II. Related Work

A. Robot Politeness

Sometimes, robots need help to accomplish their goals. There are many ways for robots to ask for help, some polite, some impolite, some effective and some ineffective [38]. In order to move through a crowded space, it is important that the robot not only alerts people of its intentions, but also does so in a socially acceptable manner. Fischer *et al.* [11] studied the perception of robot friendliness and effectiveness when a robot asks for help using acoustic signals or verbal greetings. They found that participants perceived the robot as being "more friendly" with the verbal greeting because of a preferred social framing for verbal interactions, but the effectiveness in getting assistance did not differ compared to the acoustic signal condition.

For an autonomous agent to navigate in social settings, it must exhibit acceptable manners [44]. While the importance of *social appropriateness* is generally accepted in the Human-Computer Interaction (HCI) community, it is also not well-defined. Given this, we approach social appropriateness through the lens of design, more specifically, we follow the process described by Dieter Rams [20], adding as little as possible to the robot and its actions to enable the robot to navigate through the social space. It is challenging to define specific motions that are polite for robots, since politeness and appropriateness are contextually dependent [15]. Socially appropriate navigation for robots has been studied in interactions with people and assistive technology [42], interactions between walking partners [10] and navigation through an airport [21]. From these studies, we can draw conclusions about appropriate social navigation strategies for robots, for example, not walking between two people who are engaged in conversation [42].

In crowded spaces, a clear path may not exist between the robot's current position and the goal. Here, path planning must also include the option of encouraging people to move out of the way so that a clear path exists to the goal. Social navigation becomes a higher-dimensional problem where the moving agent must consider re-planning the path to follow and also encourage specific humans to move out of the way.

A robot potentially has many different strategies to attract the attention of people in the space to ask them to move. But which strategies are most effective while still being socially acceptable? How should a robot gain the attention of bystanders and how should it communicate its intent to have them move out of the way?

B. Gaining Attention

Attention-seeking behavior in human-human interactions are frequently multi-modal. Non-verbal [34] and proximal cues [14] can be used to interrupt conversation or draw attention, but such cues do not specifically indicate the robot's intent. Touching combined with an audio cue [19] or both visual and audio cues [44] are common combinations. Previous studies have demonstrated that the robot's eye gaze, blinking, and head turns are effective tools for attracting and controlling human attention if the robot is within the field of view of the subject [29]. Other researchers have found that robot gaze does not reflexively catch people's attention, more so than other directional information [1]. Bruce *et al.* [5] explored multiple modalities of expression, using attentive motion and the use of an animated face on a robot to gain attention. They found that a combination of modalities resulted in the most compelling behavior. Imai *et al.* [16] achieved joint attention between a human and a robot using eye contact, head direction, hand gestures, and relevant utterances in the context of looking at a poster together.

The importance of gaining attention, or regaining it when lost, has been documented by authors in both the HRI

Code	Definition	Examples
Event	Begins when the robot uses a modality because a person is in the way until either the human is out of the way or the modality repeats.	Stops moving because a person or chair is in the way until a person moves out of the way or a chair is moved out of the way.
Attention	Looks at the robot.	Eye gaze, head turns towards the robot.
Constructive engagement	Active motor or verbal behavior in response to the robot [32], [23].	Touches, points, waves, kicks, gestures to pass, talks about robot, approaches robot to interact.
Human moves out of path	Begins once robot moves after stopping.	Robot moves because human rotates/walks/steps away to the side of the path.
Robot moves out of path	Begins once robot moves and ends when human's feet are no longer in view or have stopped moving in view.	Rotates and goes off path to bypass a human or obstacle.

TABLE I
QUALITATIVE CODING SCHEMA: CODE, DEFINITIONS, AND EXAMPLES

and human psychology literature. Szafir et al. [40] used physiological correlation to show that attention-seeking robot behaviors improved measures associated with positive interaction, such as recall. For humans, shifting gaze from one stimuli to another indicates a shift in attention, a method often used in HRI research [29], [43] and which we used in our video coding schema. Most notably, the bulk of the existing literature addresses the acquisition and retention of attention when the robot is visible to a person. The problem of gaining attention from behind, although not completely neglected, has been relatively underappreciated, especially in social contexts.

C. *Communicating Intent*

A clear demonstration of intent is important when warning people about what we are about to do next [41], [36], which is important for avoiding conflicts [22]. Between people, intention cues are often subtle and directed; for example, offering to hold a door open by looking at another person while holding the door open. Humans naturally empathize with other peoples' posture and actions, allowing us to understand intentions in social situations. Purely capturing attention from other social actors is not sufficient to achieve the goal of moving through a crowded space. There also needs to be communication of the intent to move past. Baraka *et al.* [3] found that using different color light modalities (i.e. red to signal an obstructed path) was successful to indicate the robot's intentions. In a human-to-human interaction, communicating intent is often accomplished using visual, auditory, and motion cues [22]. For humans, a forward-directed gaze and a confident gait can implicitly communicate intent. Typically, robots do not naturally display traits that evoke empathy, and while in practice the use of similar traits is possible in robots [8], [45], their effective use for communicating intent is not easily achieved. As an alternative to implicit communication, robots must rely on more explicit communication of intent, which is the method we use here.

III. METHOD

We conducted an in-the-wild study using haptic, visual and audio modalities to investigate capturing attention and communicating intent in a social setting. Data was collected at social events over a time span of 12 days. All participants signed legally-approved consent declarations which did not divulge the mechanisms to be examined in the experiment.

Conditions 1 (no modality) and 2 (haptic only) were collected 12 days apart from Conditions 3 (haptic and visual only) and 4 (haptic, visual and audio). See Section III-E for details on the manipulations used in each condition. Between Conditions 1 & 2, and 3 & 4, there was a one hour technical presentation on topics distinct from this study. All conditions used the same robot. All WoZ tele-operations were conducted from a fixed, hidden location. Conditions 1, 3, and 4 were tele-operated by the same researcher with a different researcher conducting Condition 2.

The experimental space was monitored using robot-mounted and environment-mounted cameras to simultaneously record the event from multiple vantage points. Four GoPro cameras were mounted on the walls/ceiling of the social interaction space while one GoPro camera was mounted on the back of the robot. The robot had two built in cameras: one looking at the floor space in front of it and one looking forwards. The tele-operator used these built in cameras to navigate.

A. *Participants*

Participants were recruited through convenience sampling: e-mail lists and word of mouth. There were a total of N = 25 unique participants across the social events, of which N = 23 adults (4 female) completed the questionnaire afterwards, while N = 19 were behaviorally coded. The others were bystanders to the robot at the parties and did not interact with the robot. Each of the 25 participants provided demographic information including sex, age, education level and experience with robots. The participant ages ranged from 21 to 45 (M = 29, SD = 5.8) and on average, were halfway between intermediate and advanced in their experience with robots (M = 3.5, SD = 1.0), where 1 = Fundamental Awareness and 5 = Expert [30]. Bystanders could attend the event, and still disagree to have their data shared externally and did not participate in the questionnaire part of the study.

As our study was based on opportunistic interactions between robots and humans who happened to be in the robot's path, it was not possible to ensure that there were an equal number of interactions in each of the four conditions. In addition, some participants attended more than one party: N = 10 participants took part once, a further N = 10 took part twice, N = 2 participants took part three times and N = 1 participant took part in all four parties. This resulted in the following numbers of participants in each condition, for video coding: C1, N = 9; in C2, N = 2; in C3, N = 11; in C4, N = 4; and for the questionnaires C1, N = 10; C2, N = 10; C3, N = 13; C4, N = 6.

There is a variable number of times that each participant attended a condition. N = 12 individuals were coded in one condition, N = 5 individuals were coded in two conditions, and N = 1 individual was behaviorally coded in three conditions. In terms of the questionnaires, one individual took the questionnaire in four conditions, two in three conditions, 10 in two conditions, and 9 in one condition.

B. Behavioral Measures

Each condition was monitored by six time-synchronized cameras (two on the robot and four GoPro's on the walls). With six video data streams, 20-30 minutes of data collection per data stream and four conditions, a total of approximately 20 hours of video data were collected. The video data was coded using ELAN 5.7 software [12], with the six video files synchronized on a split screen (Figure 1). In order to validate the reliability of the video coders, both coders coded 10% of each condition, which was used to rate the inter-rater reliability Cohen‘s Kappa, κ = 0.67. Table I summarizes the video coding scheme used by the researchers, their definitions and examples.

The video sequence was broken down into a sequence of events. An *event* is defined as beginning when the robot stops because a person is in the way of the robot's path until the path is clear and the robot can move. Within the events, we measured the following properties of each interaction:

- **Interaction time** The time between when the robot prompts a modality and the robot begins to move. This value was averaged over interactions to compute average interaction time for each condition.
- **Participant moving outcome** Interactions can be a success, in which the participant moves out of the way so that the robot can continue along its path, or a failure, in which the participant does not move out of the way.
- **Constructive engagement** A constructive engagement occurs if a participant engages with the robot through an active motion or verbal behavior in response to the robot [32], [23].

C. Subjective Measures

Since there is a lack of commonly agreed upon HRI questionnaires, researchers tend to modify existing questionnaires for relevance to their specific problem or task [37]. To measure the participant's perceptions of the robot's social behavior, we utilized the Interpersonal Dominance Scale [7] (used in HRI studies [33], [2], [6]) which measures perception of an actor's behavior along five dimensions: poise, persuasion, conversational control, panache, and self-assurance. Responses to individual questions were grouped into the five dimensions of the Interpersonal Dominance Scale. These five dimensions and their respective questions are shown in Table III. We modified the 32-item scale to be relevant in our context. This included adapting the questions to have the robot as the focus, such as "This robot seemed present to people in the room" and "This robot moved with self-confidence when interacting with others." The participants could indicate their agreement by selecting an answer from a Likert scale where 1 = strongly disagree and 7 = strongly agree. The individual questions are described in Table II. Three of the questions, marked with an R, were reverse scored for the data analysis. Individual item scores were averaged by participant to form the aggregate scores given in Table III. Additionally, we asked the question used in Shen *et al.* [37], "How well would you rate the robot in terms of social interaction?" in which participants could indicate their responses by selecting on a 7 point Likert scale, from "not good" to "very good."

Fig. 2. The tele-operated Omni Robot was instrumented with a fan, mounted just below the screen, speakers, a screen, as well as a visible set of eyes.

D. Robot System

We used a modified OhmniLabs Telepresence Robot, Developer Edition [31] in this study (Figure 2). The OhmniLabs robot is 1.4m tall, with a 22 x 14cm screen and a 37cm x 40cm base. It has a maximum speed of 1m/s and weighs 9 kg. For tele-operation, the robot has a front view and a top-down view camera.

The robot was augmented with three different interaction mechanisms:

- **Haptic mechanism** The robot was augmented with a computer controlled USB-powered fan that could provide an air-delivered touch to participants near the robot. We used the interaction engine from Matelaro *et al.* [28] to control a relay that provided power to the fan. The control was operated by the WoZ tele-operator through buttons on a web-page. The air fan was chosen as the delivery mechanism for a haptic touch given its relative safety compared to physical touch.

- **Audio mechanism** Speakers mounted on the robot were used to provide audio cues for two seconds. Audio cues were provided via a pre-recorded audio file stored on the robot. The audio cue was created by processing/stylizing the voice recording of a researcher who was articulating how the robot should sound. Using python bindings [18] for PRAAT [4] we generated a tracking sinewave of the f0 (fundamental frequency) that was used in combination with the original audio (using a vocoder) to generate the final sound. The audio file was played at a constant level that could be heard 2m from the robot. The audio cue was a higher pitched abstracted version of "excuse me."
- **Visual mechanism** A visual webcam from [17] was used to intermittently display the text message "Can you move off the line so I can get through?" on a screen mounted on the robot.

E. Manipulations

The robot relied on three basic interaction modalities; haptic, audio and visual to cue attention in a crowded space. These basic modalities were integrated in the following test conditions.

- **Condition 1: No Modality** When the robot was blocked, it waited until the person moves out of the way.
- **Condition 2: Haptic** This event involved turning the fan on for five seconds and then turning it off for five seconds. This pattern was repeated a maximum of four times.
- **Condition 3: Haptic and visual** The wizard intermittently displayed the text message "Can you move off the line so I can get through?" while the haptic cue was provided. An event involved playing this text message for 10 seconds. When the message was not displayed, a blank screen was shown. For the first five seconds of this display the fan turned on providing a haptic stimulus, and then turned off for the last five seconds. When this sequence was repeated, the visual display stayed on. The sequence was repeated a maximum of four times.
- **Condition 4: Haptic, visual and audio** Synchronized with the start of the audio cue, the haptic and visual display described under Condition 3 were provided. This sequence was presented up to four times.

After five minutes, if the robot was not able to compel the person to move out of the way, either because it could not attract the person's attention or because the person did not understand that they needed to move, the interaction was deemed a "failure." In such a case, if possible, the robot moved around the person and the robot continued on its way to the next interaction.

Label	Items
Focused	The robot is focused on its goal
Presence	The robot seems present to people in the room
Confidence	The robot moves with self-confidence when interacting with others
Nervous (R)	The robot often moves with nervousness
Expressive	The robot movements are very expressive
Dramatic	The robot has a dramatic way of interacting
Relaxed	The robot is usually relaxed and at ease
Draws Others	The robot has a way of interacting that draws others to him/her
Task-Oriented	The robot remains task-oriented during situations
Poise	The robot shows a lot of poise during interactions
Smooth (R)	The robot is not very smooth in communication
Impatient (R)	The robot is often impatient
Dictates others	The robot's movements seem to dictate those of others
Memorable	The robot's interactions were memorable

TABLE II

OUR MODIFIED INTERPERSONAL DOMINANCE SCALE QUESTIONNAIRE ITEMS AND LABELS. R = REVERSED. ANSWERS WERE SCORED ON A 7 POINT LIKERT SCALED FROM "STRONGLY DISAGREE" TO "STRONGLY AGREE."

F. Procedure

Each data collection session began with an informed consent process. Following this, we invited people to socialize and eat. For each condition, the robot was tele-operated at speeds up to 1m/s on the pre-determined path for 20-30 minutes. The robot followed a 31m continuous path that was marked on the floor using tape; this ensured consistency of the robot's path across conditions. The robot tele-operator controlled the robot to follow this path during the social events as well as controlling the attention seeking mechanisms. Figure 1 provides views from the camera of the test environment and of the robot itself. The tele-operator followed a guide based on social appropriateness in movement of motorized wheel chairs [43]. In particular, the robot stopped on the route if (a) two people were having a conversation across the robot's foreseen path or (b) someone was standing or moving on the path. The tele-operator prompted the robot's condition-specific modality. In order to move again, the tele-operator must see a clear path.

After the completion of the route 3-4 times, the robot departed the scene. The actual number of cycles of the route depended upon the end of the social event, robot connection, or the robot having to wait for too long for a participant to move. At the end of each condition participants completed the questionnaire (Table II) that probed their experiences while interacting with the robot at the party.

IV. RESULTS

A. Behavioral Results

Condition 1, No modality (C1) is essentially a control condition within which the robot makes no effort to engage with humans in the space and simply waits until the space is

(a) **C2**
11 events

(b) **C3**
64 events

(c) **C4**
21 events

Fig. 3. Interaction success rates by condition. (a) C2, haptic only, (b) C3, haptic and visual only, (c) C4, haptic, audio and visual cues. The legends also shows the number of interaction events per condition.

Dimension	Averaged Items
Self-Assurance	Focused, Nervous
Panache	Presence, Nervous, Expressive, Dramatic, Task-Oriented, Impatient, Memorable
Conversational-Control	Presence, Confidence, Task-Oriented
Poise	Confidence, Nervous, Relaxed, Draws others, Task-Oriented, Poise, Smooth, Impatient
Influence	Focused, Expressive, Dictates others

TABLE III
THE FIVE DIMENSIONS OF THE INTERPERSONAL DOMINANCE SCALE [7].

clear to move. Thus any successful interaction of the robot with users in the space are in a sense accidental. Although there were examples in which the robot was successful in having people move out of the way, a much more common outcome was the experimenter labelling this motion a failure and tele-operating the robot around the obstruction. Given that there was a lack of prompted interactions from the robot, we do not report behavioral results for C1. C1 participants did complete the questionnaire following their interaction with the robot, and subjective results for C1 are reported below.

Figure 3 shows interaction success rates for conditions C2–C4. If after an interaction, a person moved out of the way, it was counted as a success. If after the interaction, a person did not move out of the way, it was counted as a failure. There was a trend for the success rate of these interactions to increase, rising from 18% for C2 to 57% in C4.

The graph labelled, “Success Only”, in Figure 4 plots the mean duration of interactions and the mean duration of successful interactions for Condition 2 through Condition 4. As the number of interaction modalities increased, there was a trend for the duration of the interaction to decrease.

We found that many participants were treating the robot as a social actor. For example, P23 exclaimed, “Hey Googly eyed robot” in the middle of his own conversation. Also, P20 gestured towards the path for the robot to move. This could imply that the participant assumed the robot had an understanding of what that gesture meant. In Condition 2, P15, exclaimed, ”So I have to move because of this?”, while gesturing towards the robot. This could indicate that the participant understood that this was the robot’s intent.

B. Subjective Results

Individual participant responses to the questionnaires that make up each questionnaire dimension of the Interpersonal Dominance Scale were averaged per condition. Figure 5 plots the average results along with standard errors. The nature of the subject pool associated with the in-the-wild structure of the experiment conducted precludes the use of strong statistical tests to compare results across conditions. The following trends can be observed from the graphs.

- There is an increasing perceived self assurance, panache, conversational control, poise and influence of the robot as the complexity of the manipulations increased.
- The No Modality (C1) condition results in the lowest score across all five measures while the haptic, visual and audio (C4) condition results in the highest score across all conditions.

V. DISCUSSION

For a robot to navigate in congested social environments, it must be able to capture the attention of individuals in the space and communicate its intention to move through that space. This paper considers mechanisms that a robot can use to capture human attention and communicate intent while being both effective and socially appropriate. Here, we examined how well haptic, visual and audio cues might provide these tools. We found that as the number of interaction modalities increased, the interactions tended to be shorter and more effective. Additionally, participant’s perception of the robot showed increased self-assurance, panache, conversational control and influence over the interaction. Adding

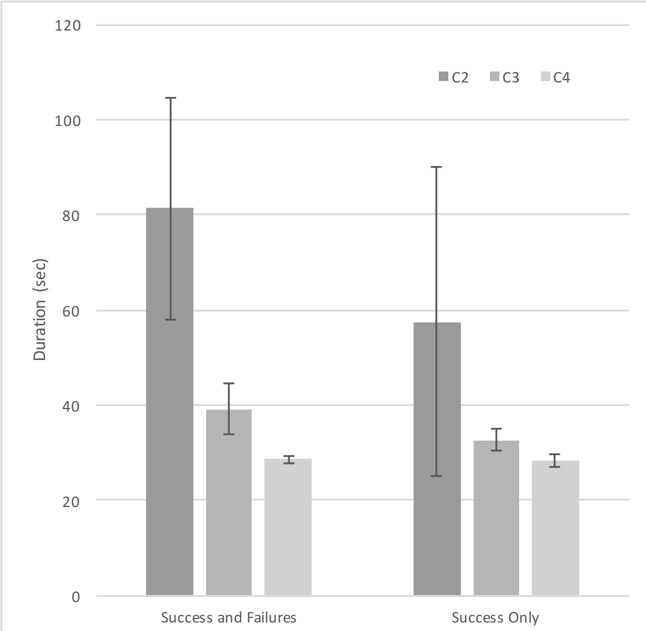


Fig. 4. Left: average wait times aggregated across failing and successful interactions, Right: average wait times across the events that had a successful outcome. Error bars show standard errors.

additional modalities not only resulted in more effective and efficient interactions, it also potentially made the robot appear more socially effective.

The modified version of the Interpersonal Dominance Scale provided insights into the participants' perceptions of the robot in a social space. The results indicate that the introduction of different modalities changed the participant's perceptions in each of the five dimensions. Participants' ratings of conversational control and influence increased with condition 4 rated highest. The dimensions of self-assurance and panache relate more to the presentation of the robot in the social space and these also increased across conditions. The results also indicate that the modified Interpersonal Dominance measure tended to be sensitive to the effects of the different modalities. This measure could be used in further studies comparing different means of attracting attention and conveying intention.

To understand if the participants interpreted intent, and to inform the Likert scale questionnaire, we asked, "If at all, what was the robot trying to communicate to you?" Participants indicated that they either failed to understand what the robot wanted to accomplish (i.e. move through the location they were standing), or that the robot needed them to move out of the way to accomplish this. In future work, we would like to qualitatively analyze these responses more thoroughly.

The modalities provided here were delivered in a synchronized manner. Natural human to human interactions often includes a temporal sequencing component that we did not investigate here. In future studies, it will be interesting to explore if temporal sequencing of haptic and audio cues are more nuanced and effective in getting attention.

The haptic cue was directed wind energy created by a fan on the robot. This provided a gentle physical stimulus to anybody in front of the robot up to a distance of 1.8m. To our knowledge, this is the first use of this type of cue in a robot navigation context, and it proved valuable in discreetly attracting the attention of party-goers who were initially facing away from the robot. While it is outside the scope of this paper, we also examined a range of alternative fan arrangements to evaluate the tradeoffs between fan speed, size, etc. This particular arrangement provides good range, a measure of directionality, and low noise. Notably, for some fan arrangements there is the risk of distracting incidental bystanders due lack of focus of the wind or excessive sound production. These engineering issues are important in practice, but incidental to the human-interaction focus reported here.

The visual modality used here relied on a text display. For the participant to understand the robot's intent they must turn to view the display. Although this was a particularly appropriate mechanism for this study, it is perhaps not the most effective mechanism when used without another cue. For example, an audio cue might both provide an "excuse me" message as well as a "May I get through" message as part of a single utterance. However, relying on haptic and visual modalities, may be especially useful in a noisy environments. In future studies, we would also like to utilize the social framing used in Srinivasan *et al.* [38] for the words, which was effective in changing the perception of friendliness.

As the study did not explore all possible combinations of different modalities, we can make inferences on which modalities are effective and how they may interact. For a robot to be effective in navigating the space, it must not only be able to capture attention, it must also signal intent. Here we choose haptic and audio cues, two modalities that are specifically appropriate for capturing attention, based on the literature, that were likely to be effective in terms of capturing attention. We also chose a visual signal that was explicitly chosen to signal intent. Although there are certainly other possible cues that could be used to provide these signals, we found these cues to be quite effective in their specific roles.

The results showed that participants in interactions under conditions with more dimensional modalities found the robot to be more dominant in a social situation than a robot that used a lower dimensional interaction modality. Time and resources did not permit conditions to be conducted for all possible modality combinations. It would be interesting to explore how different modalities individually performed in this space. For example, is communication of intent and social effectiveness perceived if multiple modalities are used? Or, is one sufficiently rich mechanism for capturing intent enough? These are important questions for future work although without a sufficiently rich interaction mechanism it becomes difficult in an uncontrolled setting to establish any substantive interaction.

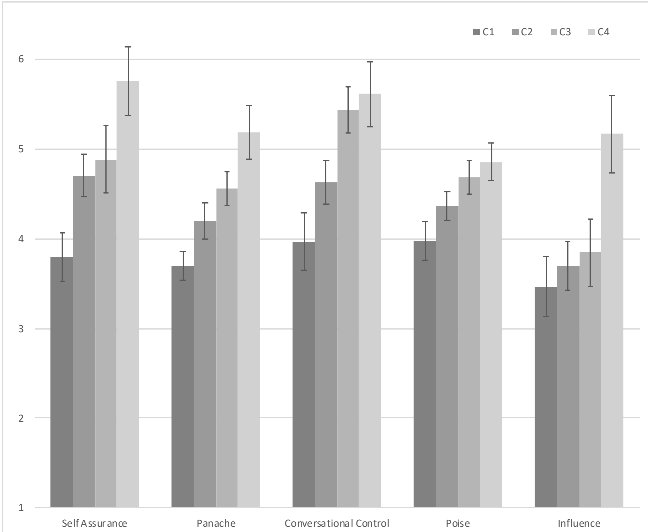


Fig. 5. Interpersonal dominance scale results. Means are plotted along with standard errors on a scale from 1 = strongly disagree to 7 = strongly agree to the robot exhibiting the corresponding social behavior.

A. Limitations

Like many in-the-wild studies, we struggled with maintaining a completely within or between-subjects study design. While we invited the same group of people to this weekly event, the same group did not come to each event. A study with access to a larger naïve participant population that consistently attended, would have permitted a more controlled experiment and thus a more sophisticated data analysis. The nature of the study reported here prevents the use of strong statistical tools to analyze formally the results plotted in Figure 5. Despite that, the underlying trend visible in Figure 5 seems to suggest a consistent result.

It becomes difficult to measure the effectiveness of displacing people when participants are testing or playing with the robot. For example, in C4, a number of the interactions included participants getting in the way of the robot and moving out of the way on purpose.

As in any real robot system, we dealt with several hardware issues, including a poor wireless data connection, in which the robot would stop until it re-connected to its wireless base station. Also, despite having six camera views, the video coders could not always see everything. Additionally, as with many tele-operation studies [25], [39], [27], lag can be an issue. Lastly, the tele-operators prompted the fan in proximity in which the participants could feel the fan (by adding visual tools to the wizarding interface), but didn't always make that necessary proximity, so the feeling of the fan is not guaranteed. In the future, we would like to control consistent proximity more effectively.

It is possible that the novelty effect [13] caused more constructive engagement (i.e., testing the robot, putting a chair in front of the robot's path, poking at the robot, looking for sensors, or poking to understand what prompts the fan) than would be found after a longer-term deployment.

The fixed path followed by the robot was not especially social, since moving along a fixed path may not be the most considerate behavior for the robot. We felt that being consistent about the path throughout all conditions outweighed this social benefit. In the future, we would like to explore more flexible paths with a consistent general direction.

Finally, we also acknowledge that the Wizard-of-Oz method can introduce variability: the wizard could change unconsciously over time. While we developed detailed instructions for the remote operator, including timing rules, it is often impossible to operate the robot with exactly the same intervention times, as in other HRI studies [39]. To mitigate this, we automated the succession of some modalities after the wizard prompted the modality. For example, in condition 4, turning on the fan also automatically turned on the audio cue.

VI. Conclusion

In this paper, we examined mechanisms to allow a robot to move through a social gathering effectively and socially. We employed a unique haptic cue – blowing wind – in combination with a visual stimulus and audio, to attract the attention of people when they were facing away. These cues proved to be an efficient way to compel people to make way for the robot in a natural social setting. Additionally, we used a modified version of the Interpersonal Dominance Scale to assess social effectiveness in the space and found that the robot tended to be more socially effective in the space when more modalities were prompted.

Acknowledgments

Thank you to Yi Tian Xu and Francois Hogan for helping with copy editing and to Sandeep Manjanna for helping us in the early stages with the study design. Thanks to Samsung Artificial Intelligence Center, Montreal for supporting this research.

References

[1] H. Admoni, C. Bank, J. Tan, M. Toneva, and B. Scassellati, "Robot gaze does not reflexively cue human attention," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 33, Boston, MA, 2011.

[2] V. Ahumada-Newhart and J. S. Olson, "Going to school on a robot: Robot and user interface design features that matter," *ACM Trans. Comput.-Hum. Interact.*, vol. 26, no. 4, pp. 25:1–25:28, June 2019. [Online]. Available: http://doi.acm.org/10.1145/3325210

[3] K. Baraka, S. Rosenthal, and M. Veloso, "Enhancing human understanding of a mobile robot's state and actions using expressive lights," in *Proceedings 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. New York, NY: IEEE, 2016, pp. 652–657.

[4] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," Version 6.0.37, retrieved 3 February 2018 http://www.praat.org/, 2018.

[5] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human-robot interaction," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (ICRA)*, vol. 4, Washington, DC, May 2002, pp. 4138–4142 vol.4.

[6] J. K. Burgoon, J. A. Bonito, B. Bengtsson, C. Cederberg, M. Lundeberg, and L. Allspach, "Interactivity in human–computer interaction: A study of credibility, understanding, and influence," *Computers in Human Behavior*, vol. 16, no. 6, pp. 553–574, 2000.

[7] J. K. Burgoon, M. L. Johnson, and P. T. Koch, "The nature and measurement of interpersonal dominance," *Communications Monographs*, vol. 65, no. 4, pp. 308–335, 1998.

[8] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler, and J. Forlizzi, "Using spatial and temporal contrast for fluent robot-human handovers," in *Proceedings 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Lausanne, Switzerland, March 2011, pp. 489–496.

[9] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of oz studies - why and how," *Knowledge-Based Systems*, vol. 6, no. 4, pp. 258–266, 1993.

[10] G. Ferrer, A. G. Zulueta, F. H. Cotarelo, and A. Sanfeliu, "Robot social-aware navigation framework to accompany people walking side-by-side," *Autonomous Robots*, vol. 41, no. 4, pp. 775–793, 2017.

[11] K. Fischer, B. Soto, C. Pantofaru, and L. Takayama, "Initiating interactions in order to get help: Effects of social framing on people's responses to robots' requests for assistance," in *Proceedings 23rd IEEE International Symposium on Robot and Human Interactive Communication*, Edinburgh, Scotland, Aug 2014, pp. 999–1005.

[12] N. M. P. I. for Psycholinguistics, *ELAN (Version 5.2) [Computer software]*, Accessed: 2020-01-02. [Online]. Available: Retrievedfromhttps://tla.mpi.nl/tools/tla-tools/elan/

[13] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. C. Schultz, *et al.*, "Designing robots for long-term social interaction," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2005, pp. 1338–1343.

[14] K. Hayashi, M. Shiomi, T. Kanda, and N. Hagita, "Friendly patrolling: A model of natural encounters," in *Robotics: Science and Systems VII*. Sydney, Australia: MIT Press, 2012, p. 121.

[15] G. Hoffman and W. Ju, "Designing robots with movement in mind," *Journal of Human-Robot Interaction*, vol. 3, no. 1, pp. 91–122, 2014.

[16] M. Imai, T. Ono, and H. Ishiguro, "Physical relation and expression: Joint attention for human-robot interaction," *IEEE Transactions on Industrial Electronics*, vol. 50, no. 4, pp. 636–643, 2003.

[17] V. M. Inc., "Manycam," (Accessed January 2, 2020), https://manycam.com.

[18] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing Parselmouth: A Python interface to Praat," *Journal of Phonetics*, vol. 71, pp. 1–15, 2018.

[19] S. E. Jones and A. E. Yarbrough, "A naturalistic study of the meanings of touch," *Communications Monograph*, vol. 52, no. 1, pp. 19–56, 1985.

[20] C. W. D. Jong, Ed., *Dieter Rams: Ten Principles for Good Design*. New Yorl, NY: Prestel, 2017.

[21] M. Joosse and V. Evers, "A guide robot at the airport: First impressions," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 149–150. [Online]. Available: http://doi.acm.org/10.1145/3029798.3038389

[22] W. Ju, "The design of implicit interactions," *Synthesis Lectures on Human-Centered Informatics*, vol. 8, no. 2, pp. 1–93, 2015.

[23] K. S. Judge, C. J. Camp, and S. Orsulic-Jeras, "Use of montessori-based activities for clients with dementia in adult day care: Effects on engagement," *American Journal of Alzheimer's Disease*, vol. 15, no. 1, pp. 42–46, 2000.

[24] M. Jung and P. Hinds, "Robots in the Wild: A Time for More Robust Theories of Human-Robot Interaction," *ACM Transactions on Human-Robot Interaction*, vol. 7, 2018.

[25] D. B. Kaber, J. M. Riley, R. Zhou, and J. Draper, "Effects of visual interface design, and control mode and latency on performance, telepresence and workload in a teleoperation task," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44. Santa Monica, CA: SAGE Publications, 2000, pp. 503–506.

[26] J. F. Kelley, "An iterative design methodology for user-friendly natural language office information applications," *ACM Transactions Information Systems*, vol. 2, no. 1, pp. 26–41, Jan. 1984. [Online]. Available: http://doi.acm.org/10.1145/357417.357420

[27] S. I. MacKenzie and C. Ware, "Lag as a determinant of human performance in interactive systems," in *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. Amsterdam, The Netherlands: ACM, 1993, pp. 488–493.

[28] N. Martelaro, M. Shiloh, and W. Ju, "The interaction engine: Tools for prototyping connected devices," in *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '16. New York, NY, USA: ACM, 2016, pp. 762–765. [Online]. Available: http://doi.acm.org/10.1145/2839462.2854117

[29] M. Moshiul Hoque, T. Onuki, Y. Kobayashi, and Y. Kuno, "Effect of robot's gaze behaviors for attracting and controlling human attention," *Advanced Robotics*, vol. 27, no. 11, pp. 813–829, 2013.

[30] U. N. I. of Health (NIH), "Competencies proficiency scale," https://hr.nih.gov/working-nih/competencies/competencies-proficiency-scale, Accessed: 2019-07-20.

[31] Ohmnilabs, *Ohmni Supercam Telepresence Robot*, Accessed: 2019-08-08. [Online]. Available: http://ohmnilabs.com/telepresence/

[32] S. Orsulic-Jeras, K. S. Judge, and C. J. Camp, "Montessori-based activities for long-term care residents with advanced dementia: effects on engagement and affect." *The Gerontologist*, vol. 40, no. 1, pp. 107–111, 2000.

[33] I. Rae, L. Takayama, and B. Mutlu, "The influence of height in robot-mediated communication," in *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*. Tokyo, Japan: IEEE Press, 2013, pp. 1–8.

[34] P. Saulnier, E. Sharlin, and S. Greenberg, "Exploring minimal nonverbal interruption in HRI," in *2011 IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Atlanta, GA: IEEE, July 2011, pp. 79–86.

[35] S. Savanovic, M. P. Michalowski, and R. Simmons, "Robots in the wild: observing human-robot social interaction outside the lab," in *Proceedings of the IEEE international Workshop on Advanced Motion Control*, Istanbul, Turkey, 2006.

[36] A. E. Scheflen, *Body Language and the Social Order; Communication as Behavioral Control*. Englewood Cliffs, NJ: Prentice Hall, 1972.

[37] Q. Shen, K. Dautenhahn, J. Saunders, and H. Kose, "Can real-time, adaptive human–robot motor coordination improve humans overall perception of a robot?" *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 1, pp. 52–64, 2015.

[38] V. Srinivasan and L. Takayama, "Help me please: Robot politeness strategies for soliciting help from humans," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: ACM, 2016, pp. 4945–4955. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858217

[39] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common metrics for human-robot interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*. Montreal, Canada: ACM, 2006, pp. 33–40.

[40] D. Szafir and B. Mutlu, "Pay attention!: designing adaptive agents that monitor and improve user engagement," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Austin, TX: ACM, 2012, pp. 11–20.

[41] R. M. Uchanski, "Clear speech," *The Handbook of Speech Perception*, vol. 2, p. 708, 2005.

[42] D. Vasquez, P. Stein, J. Rios-Martinez, A. Escobedo, A. Spalanzani, and C. Laugier, "Human aware navigation for assistive robotics," in *Experimental Robotics: The 13th International Symposium on Experimental Robotics*, J. P. Desai, G. Dudek, O. Khatib, and V. Kumar, Eds. Heidelberg: Springer International Publishing, 2013, pp. 449–462. [Online]. Available: https://doi.org/10.1007/978-3-319-00065-7_31

[43] M. Vázquez, E. J. Carter, B. McDorman, J. Forlizzi, A. Steinfeld, and S. E. Hudson, "Towards robot autonomy in group conversations: Understanding the effects of body orientation and gaze," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 42–52. [Online]. Available: http://doi.acm.org/10.1145/2909824.3020207

[44] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. Te Boekhorst, and K. L. Koay, "Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion," *Autonomous Robots*, vol. 24, no. 2, pp. 159–178, 2008.

[45] S. Yang, B. K. Mok, D. Sirkin, H. P. Ive, R. Maheshwari, K. Fischer, and W. Ju, "Experiences developing socially acceptable interactions for a robotic trash barrel," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Kobe, Japan: IEEE, Aug 2015, pp. 277–284.

[46] S. Yilmazyildiz, R. Read, T. Belpeame, and W. Verhelst, "Review of semantic-free utterances in social human–robot interaction," *International Journal of Human-Computer Interaction*, vol. 32, no. 1, pp. 63–85, 2016.