# Enthusiastic Robots make Better Contact

Saad, Elie; Broekens, Joost; Neerincx, Mark; Hindriks, Koen

**DOI**
10.1109/IROS40897.2019.8967950

**Publication date**
2019

**Document Version**
Accepted author manuscript

**Published in**
2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2019

**Citation (APA)**
Saad, E., Broekens, J., Neerincx, M., & Hindriks, K. (2019). Enthusiastic Robots make Better Contact. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2019* (pp. 1094-1100). Article 8967950 (IEEE International Conference on Intelligent Robots and Systems). IEEE. https://doi.org/10.1109/IROS40897.2019.8967950

*Published in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) Macau, China, Nov. 4-8, 2019*

# Enthusiastic Robots Make Better Contact

Elie Saad[1], Joost Broekens[2], Mark A. Neerincx[1,3] and Koen V. Hindriks[4]

***Abstract*— This paper presents the design and evaluation of human-like welcoming behaviors for a humanoid robot to draw the attention of passersby by following a three-step model: (1) selecting a target (person) to engage, (2) executing behaviors to draw the target's attention, and (3) monitoring the attentive response. A computer vision algorithm was developed to select the person, start the behaviors and monitor the response automatically. To vary the robot's enthusiasm when engaging passersby, a *waving* gesture was designed as basic welcoming behavioral element, which could be successively combined with an *utterance* and an *approach* movement. This way, three levels of enthusiasm were implemented: Mild (waving), moderate (waving and utterance) and high (waving, utterance and approach movement).**

**The three levels of welcoming behaviors were tested with a *Pepper* robot at the entrance of a university building. We recorded data and observation sheets from several hundreds of passersby ($N = 364$) and conducted post-interviews with randomly selected passersby ($N = 28$). The level selection was done at random for each participant. The passersby indicated that they appreciated the robot at the entrance and clearly recognized its role as a welcoming robot. In addition, the robot proved to draw more attention when showing high enthusiasm (i.e., more welcoming behaviors), particularly for female passersby.**

***Index Terms*— Human Robot Interaction; Robot Behaviors; Drawing Attention; Social Robotics; Enthusiastic Robots; Greeting Model.**

## I. Introduction

Pro-actively drawing attention by a robot involves three main steps: (1) selecting a target to engage, (2) executing behaviors to draw attention, and (3) monitoring the response. Following this strategy (Fig. 1) allows a social robot to take the initiative for social interaction. This is useful due to the growing interest for integrating humanoids in our daily life (e.g., schools [1], healthcare [2], museums [3] and shopping malls [4]). To deploy robots in these contexts, it is necessary that they are equipped with social abilities to facilitate the interaction with people and improve the assistive function [5]. The long term goal is to improve the awareness of a social robot by providing it with the means to successfully select potential communication partners; engage in an appropriate way; and monitor the attentive response.

As a first step in proactively engaging someone, a robot shall assess whether that person is interested in interacting with it [6]. Identifying potential candidates (i.e., targets) depends on multiple contextual and individual factors (e.g., distance and orientation) [7], [8]. In addition, the displayed behaviors shall be relevant to the social context and easily understandable by the communication partner [9]. Timing when to start or stop an attempt (e.g., [10]) is also critical in order to successfully draw a passerby's attention and initiate engagement [11], [12]. Once the engagement is initiated, monitoring the target's response is important for deciding whether to maintain it or disengage accordingly [13].

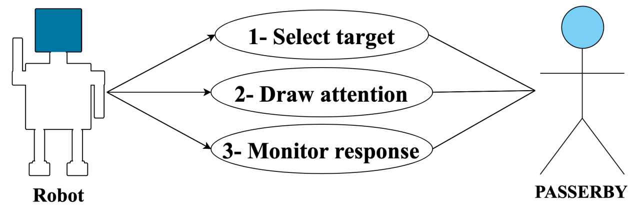


Fig. 1. Illustration of the three-step model.

In human-human interaction (HHI), multiple behavioral elements are used for drawing attention and greeting people depending on the contextual and environmental factors (e.g., greeting setting and distance) [14], [15]. These elements (e.g., waving, utterance and approach movement) are useful for expressing communicative goals [16] and conveying *enthusiasm* [17]. In addition, combining the behavioral elements leads to a variation in the intensity of the behavior and the enthusiasm (i.e., excitement) of the greeter [14].

In this research we investigate whether similar human-human behaviors can also be effectively used by a social robot for drawing attention. We designed three behaviors by varying the level of enthusiasm using a basic greeting behavior (i.e., waving) that can be combined with two other greeting behaviors [14]. We then equipped a humanoid with the capabilities to act as a welcoming robot at the entrance of a building [18] and follow our proposed three-step model (Fig. 1). The targets were selected from the passersby detected in the public zone (i.e., beyond $4m$, as defined in [19]) entering the building. The effects of the robot's behaviors on the passersby' receptive responses are observed and recorded during the study.

This paper is organized as follows. In Section II we review related work. In Section III we discuss the behavior design. In Sections IV and V, we present our hypotheses and experimental methods. In Sections VI and VII, we analyze the results and discuss the findings. In Section VIII we conclude the paper.

[1]Department of Intelligent Systems - II Group, Delft University of Technology, Delft, The Netherlands. E.Saad@tudelft.nl
[2]LIACS, Leiden University, Leiden, The Netherlands. Joost.Broekens@gmail.com
[3]The Netherlands Organization for Applied Scientific Research (TNO), Soesterberg, The Netherlands. Mark.Neerincx@tno.nl
[4]Department of Computer Science - Artificial Intelligence, Vrije Universiteit, Amsterdam, The Netherlands. K.V.Hindriks@vu.nl

## II. Related Work

As pointed out by [20], the presence of robots in public spaces is still novel and may affect how people perceive and react to them. Studies have shown that this novelty effect wears off over time and that people adapt their attitudes toward robots and adjust their expectations [21], [22]. In this research, we will analyze our observations of the passersby' behaviors to check the wearing-off of the novelty effect in a welcoming interaction setting.

In HHI, speech is naturally coordinated with gestures [23]. This coordination is useful in many parts (i.e., units) of the social interaction, including greeting and initiating engagement [14]. Research in human-robot interaction (HRI) has been investigating the effects of environmental factors (e.g., space) and robot behaviors (e.g., movement) on approaching people to initiate and maintain an interaction (e.g., [24], [10]). The proposed selection of the interaction partners (i.e., targets) was based on multiple factors, for example motion [11], space [7] and walking behavior [8]. The focus of these studies was to observe people's behavior (e.g., trajectory) and decide whether to approach them or not.

The findings from Bruce et al. [25] indicate that robot behaviors which combine facial expressions with attentive movements (i.e., turning head toward a passerby) have more influence on increasing the interest of passersby (detected within $4m$). The experiment took place in a busy corridor of a university building. Similarly, Finke et al. [11] conducted a study on attracting people's attention in a public corridor and from close distance (within $2m$). Their findings indicate that orienting the robot (i.e., turning it toward passersby) is not enough for drawing people's attention (which, according to the authors, may be due to the slowness of the robot). Other conditions (e.g., speech and gestures) were not investigated. Furthermore, Torta et al. [26] conducted a study with elderly people (62-70 years). They evaluated different modalities (e.g., waving and saying 'Hello') for drawing attention by a remotely-controlled robot standing next to the participants. Their results show that reaction time was faster when robot actions included an auditory stimulus.

In summary, research has been conducted on approaching and initiating interaction with people from close distances (within $4m$). Better results were achieved with behaviors containing auditory stimulus. However, drawing people's attention from farther distances has not been investigated yet. In addition, providing the robot with the means to monitor the attention of passersby requires further investigation. In our study, we aim for drawing the attention of passersby located in the public zone, by complementing approach movement with other behaviors (e.g., waving and speech).

## III. Behavior Design

In this section, we will present the welcoming interaction scenario, the robot capabilities and the three-step model.

### A. Scenario

In a welcoming social interaction, a host greets guests using an exchange of gestures and/or utterances [14]. In our setting, the main task of our robot (i.e., the host) is to welcome visitors entering a building. This requires that it distinguishes between people entering (i.e., proceed with the greeting) from those who are exiting the building (i.e., do not engage). With the focus on people entering, three situations may occur. The first is when a person is entering alone, the robot shall greet that person. The second occurs when people enter in groups (of two or more). In this case, only one greeting shall be executed by randomly selecting a target from the group. The third happens when people enter sequentially (i.e., quickly one after the other). In this case, the robot shall greet the first entering person. This will prevent the robot greeting from being unrecognized or rebuffed.

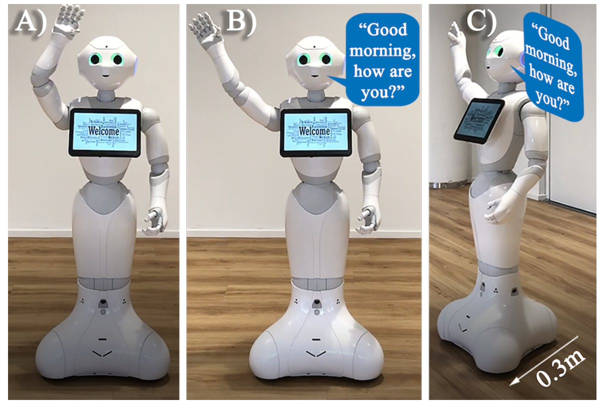


Fig. 2. Illustration of the robot *Pepper* displaying a welcome sign on its tablet and acting as a welcoming robot by executing the three behaviors with A) mild enthusiasm (one greeting element); B) moderate enthusiasm (two greeting elements); and C) high enthusiasm which combines three greeting elements (explanation in text).

### B. Robot Capabilities

In this study, we are using a humanoid robot, *Pepper*[1] (Fig. 2). When deployed at the entrance, it is expected to continuously explore the space and search for candidates to engage. To detect and extract relevant features of passersby, *Pepper*'s on-board capabilities are insufficient due to their limited range (around 4m). To extend its reach, we complemented it with state-of-the-art computer-vision techniques. As illustrated in the system architecture of Fig. 3, the frames received from *Pepper*'s front 2D camera are processed for detecting and tracking people by extracting their bounding boxes (using YOLO [27]). The boxes are then processed (using OpenPose [28]) for extracting and monitoring people's five head keypoints (two eyes, two ears and the nose) as proxy for attention.

To make our robot role obvious for passersby, its tablet is used to display a welcome sign (Fig. 2). In addition, its sensors are used to detect if someone touches its head, hands or base, to kindly ask them to stop. With the absence of background noise at the building entrance, we set the volume of its speakers to 55dB (70% of the full capacity) after testing it with a few number of passersby ($N = 11$).

[1]SoftBank Robotics, `https://www.softbankrobotics.com`

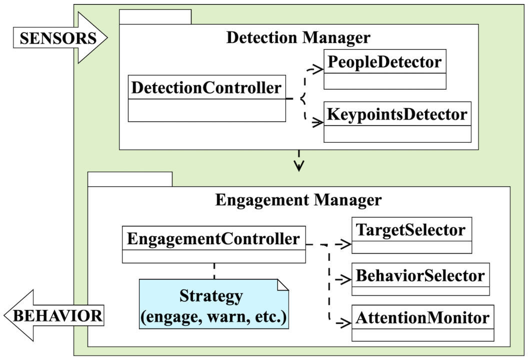


Fig. 3. High level architecture of the extended system which receives readings from robot sensors and sends a corresponding behavior to execute.

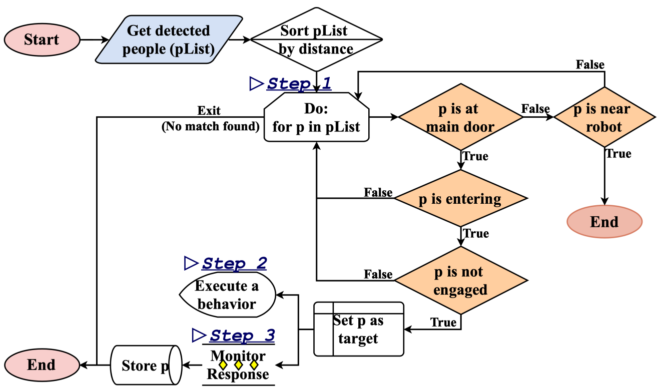


Fig. 4. Flow of the three-step model for selecting a target (Step 1); drawing attention (Step 2); and monitoring the attentive response (Step 3).

### C. Modeling the Welcoming Interaction

In this Section, we provide the details of the three-step model (Fig. 1) to manipulate the attention of passersby.

*Step 1. Target Selection:* The selection process (Fig. 4) is called whenever the robot is ready to engage a new target and people are detected at the entrance. As described in Alg. 1, the received list of detected people is first sorted by distance to the robot (line 2). Then, if the closest person has approached the robot (lines $3-4$), the subroutine exits (i.e., no need to execute a behavior); otherwise, it loops through the list (line 6) to search for a target (i.e., a match). The target has to satisfy four conditions. First, it shall not be engaged before (line 7). Second, it shall have four or more head keypoints (line 9). In this case we assume that, if the robot is in the line of sight of a person, it is expected to detect at least four out of five head keypoints. The third condition (line 11) checks if the person is located inside (i.e., the height of the bounding box is above the minimum height $h$) or outside the entrance (i.e., the height is below). The minimum height is computed based on the camera settings and the position of the robot. Finally, the fourth condition (line 13) uses the height difference ($boundingBoxHeightDiff$) $-$ computed by subtracting the previous bounding box from the current one $-$ to check if the person is entering (i.e., the difference increases) or exiting (i.e., the difference decreases).

**Algorithm 1** Select Target (*explanation in text*)

▷ Input: list of detected people ($pList$), min distance ($d$), min head-keypoints ($hkpt$), min height ($h$) and threshold ($a$)
▷ Output: target person ($p$) or null

```
 1: function FINDTARGET(pList, d, hkpt, h, a)
 2:     pList ← sort(pList, distance, increasing)
 3:     if pList[0].distance < d then
 4:         return null          ▷ closest p approached the robot
 5:     target ← null
 6:     for p in pList do
 7:         if p.isEngaged == True then
 8:             continue         ▷ skip below and keep looping
 9:         if p.headKeypoints < hkpt then
10:             continue         ▷ p is not facing the robot
11:         if p.boundingBoxHeight < h then
12:             continue         ▷ p is outside
13:         if p.boundingBoxHeightDiff > a then
14:             target ← p       ▷ p is entering
15:     return target
```

*Step 2. Attention Drawing Behaviors:* When a target is selected, our robot receives a request for executing a behavior (Fig. 4). We equipped *Pepper* with three behaviors designed using one or a combination of behavioral elements (selected from the human greetings in [14]) to vary the level of enthusiasm (Fig. 2). The first behavior, *mild* level with one element, consists of a waving gesture which is an effective non-verbal cue for attracting attention and initiating an interaction [29]. The second, *moderate* with two elements, consists of both a waving gesture combined with an utterance (i.e., the robot says: 'Good morning, how are you?'). The third, *high* with three elements, combines waving and utterance with an approach movement (i.e., the robot moves $0.3m$ forward) to reduce the distance with the selected target [24].

*Step 3. Response Monitoring:* While the robot is executing a behavior, the target's attentive response (i.e., receptiveness) is monitored in parallel (Fig. 4). The monitoring process involves tracking the distance and head keypoints of the target using the robot's cameras and sensors. The head keypoints are used for identifying passersby who are paying attention to the robot (i.e., all five keypoints are detected).

## IV. Evaluation of Behaviors

In this research, we aim for investigating the effect of welcoming behaviors (with different levels of enthusiasm) for a social robot on the attention of passersby. We formulated two hypotheses. The first is related to the novelty effect and is based on the role of our humanoid as a welcoming robot. We identified two passersby' reactions related to novelty namely approaching the robot (i.e., moving closer) and stop walking to watch it. These reactions will serve as a check for verifying the wearing-off of the novelty effect:

*H1:* The percentage of the passersby who approach or stop to watch a welcoming humanoid will decrease over time.

The second hypothesis (experimental) is based on the potential effect of the enthusiasm level of welcoming behaviors:

*H2:* A welcoming robot behavior with $high$ level of enthusiasm (i.e., which combines a waving gesture, utterance and movement), draws more attention than a behavior with $mild$ or $moderate$ level of enthusiasm.

## V. Methods

To test our hypotheses, we conducted an experiment with *Pepper* acting as a welcoming robot by following the three-step model. The experiment consisted of three phases during which we deployed *Pepper* at a university building entrance for several mornings over a period of six weeks. In phases 1 and 2, we deployed *Pepper* at the entrance in order to wear off the novelty effect (e.g., [21]) and test our system and project settings. Finally, we waited two weeks before starting phase 3. The experimental methods and procedures were approved by the human research ethics committee of the university.

### A. Experimental Design

In this study, we manipulated the enthusiasm level of the robot behavior (i.e., independent variable) by varying the number of greeting elements (Section III-C).

*1) Data Collection:* The experiment data was collected in three forms. First, the researchers collected time-stamped observation sheets to record passersby reactions (based on [11] and our observations during phases 1 and 2). Second, we collected automated data generated by the robot (e.g., entering time and head keypoints). Third, we held post-interviews conducted with randomly selected participants to obtain their feedback after being greeted by our robot.

*2) Measurements:* To compare the effect of the attention drawing behaviors, we selected different measures (i.e., dependent variables) to collect data for and analyze afterwards.

*a) Novelty effect:* To check that the novelty effect was wearing off and verify H1, we recorded the number of people who either approached the robot (i.e., moved closer to it within a distance of $1-2m$), or stopped to watch it (e.g., for $2s$ or more). This data was manually collected by the researchers using the observation sheets.

*b) Walking Speed:* To estimate the duration in the field of view of the robot, we used the entering time (when located at the main door) and leaving time (when being outside the field of view of the robot) of each participant. This data was automatically recorded by our system. The participants' walking speed was derived using (1), by dividing the distance (a constant, from the main entrance doors to the side) with the duration. The walking speed will be used to identify slow to fast people and verify H2.

$$walkingSpeed(m/s) = \frac{distance(m)}{duration(s)} \qquad (1)$$

*c) 5-Keypoint Score and Attentiveness:* First, we computed a 5-keypoint score based on what was automatically perceived by the robot (i.e., the head keypoints collected by our system and used as proxy for attention). The 5-keypoint score in itself provides already a rough attention score for each participant: 0 if no 5-keypoint was detected while being greeted by the robot; and 1 otherwise. Using the 5-keypoint data, we also computed the ratio of the time difference between the first and last detected 5-keypoint to the duration in the robot's field of view (same duration as in Section V-A.2.b). Second, based on what was perceived by the human observer (i.e., the reactions of the passersby), we assigned an attentiveness value for the passersby. If the passersby showed an interest in the robot (i.e., *look-while-walking*) they received a value of 1; otherwise, they received a value of 0 (i.e., *low or no response*). The 5-keypoint score and attentiveness measurements will be used to verify H2.

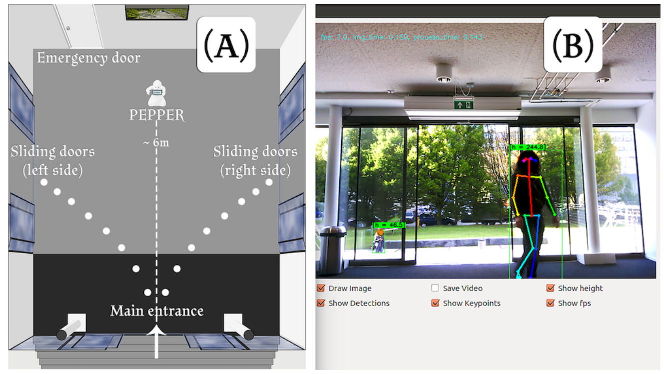


Fig. 5. A) Illustration of the building entrance. B) Part of the main user interface showing the annotated frames received from the robot's camera.

### B. Participants

Participants consisted of the passersby (new and returning) who are entering the university building. In the first two phases of our experiment, we only collected observation sheets from passersby ($N = 516$ for phase 1 and $N = 828$ for phase 2). In phase 3, we recorded data and time-stamped observation sheets from $N = 364$ passersby and conducted post-interviews with randomly selected passersby ($N = 28$).

### C. Procedure and Setup

The building entrance (Fig. 5-A) is accessed via external stairs. After passing the main doors, visitors need to traverse $5-6m$ to access the building via the left or right sides. During the experiment, we turned off the television used for displaying news and announcements (to avoid distracting people and affecting the results) and we positioned our robot in the center ($6m$ from the main doors). This position provided the robot with a central field of view and made it more visible to people entering the building (i.e., to increase its chances in drawing their attention).

The participants were exposed to a randomly selected behavior when passing by the robot. To keep track of their reactions (via the observation sheets) and to monitor the robot from a distance (via a user interface), one researcher was positioned in a room located next to the sliding doors on the right side (See Fig. 5-A). We selected this room after observing that the right side was rarely used to access

TABLE I

DISTRIBUTION OF ALL THE PASSERSBY IN PHASE 3 ($N$=364) AND THEIR REACTIONS TO THE ROBOT BEHAVIOR

| Robot Behavior | Total | Males | Females | Entered Alone | Novelty effect | Look-while-walking | Low or No response |
|---|---|---|---|---|---|---|---|
| Mild enthusiasm | 122 | 77 (63%) | 45 (37%) | 92 (75%) | 7.4% | 48.4% | 44.2% |
| Moderate enthusiasm | 123 | 91 (74%) | 32 (26%) | 84 (68%) | 9.8% | 56.1% | 34.1% |
| High enthusiasm | 119 | 98 (82%) | 21 (18%) | 86 (72%) | 8.4% | 66.4% | 25.2% |
| Combined | 364 | 266 (73%) | 98 (27%) | 262 (72%) | 8.5% | 56.9% | 34.6% |

the building (i.e., to minimize the effect of the researcher's presence).

### D. Materials

To conduct the experiment and collect the required data, we combined different materials.

*1) Questionnaire:* In the third phase of our experiment, we randomly selected passersby for a post-interview. We designed a questionnaire for the participants to indicate, on a 5-point Likert scale (the higher, the better), whether they appreciated the robot and how well they could perceive its behaviors (i.e., waving, utterance and movement). We also asked the participants to describe what the robot was doing (i.e., its assigned role).

*2) User Interface:* For the purpose of this research, we implemented a graphical interface (Fig. 5-B) to access the robot (acting autonomously) and input our observations. By means of this interface, the researchers were able to connect to the robot in order to activate its camera (e.g., to see from its field of view) and the relevant capabilities (e.g., detection techniques). They also used it to record their observations and comments via the time-stamped observation sheets.

*3) External Components:* To communicate with the robot, we setup an internal network using a wireless router (with Wi-Fi band of 5 GHz and speed of 750 Mbps). In addition, we setup a PC with a graphics card (GeForce GTX Titan X) to satisfy the requirements of the detection techniques (Section III-B). The PC was connected to the network via Ethernet cable and the robot via wireless. To improve the system performance, we used parallel threading for receiving and processing the frames from the robot's front camera (resolution set to 640x480 pixels).

## VI. RESULTS AND ANALYSIS

The distribution of the passersby (phase 3) is summarized in Table I. Around 72% entered alone and the rest entered in groups of two or more. The number of males ($N = 266$) and females ($N = 98$) was not evenly distributed between the levels of enthusiasm, which will be taken into consideration (as a control variable) during the statistical tests.

The results from the post-interviews show that 26 out of 28 participants were able to recognize the role of *Pepper* as a welcoming robot (two participants considered it as a means for displaying information and tracking building visitors). Answers on questions about noticing *Pepper* ($Mean = 4.46$, $SD = 0.74$) and appreciating it at the entrance ($Mean = 3.61$, $SD = 1.20$) suggest that Pepper's presence was well-received. Answers on questions about noticing the $waving$

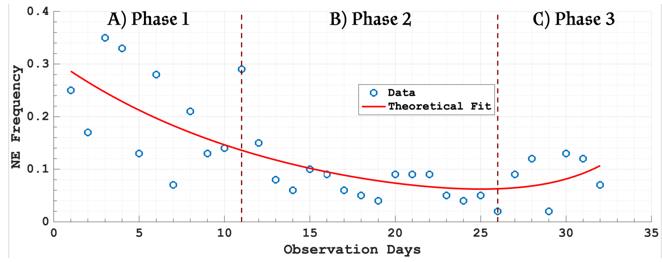


Fig. 6. Frequency of the novelty effect (NE) behaviors and trend over time, during the three phases of the experiment.

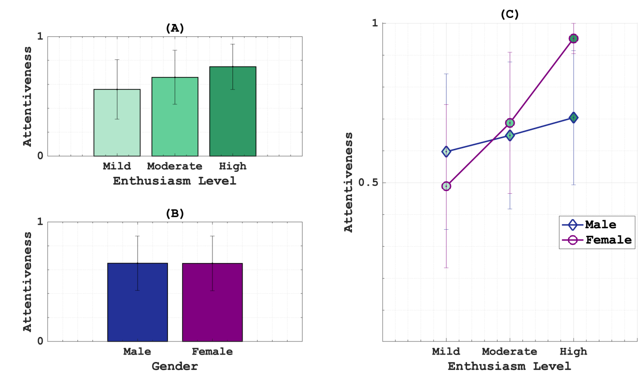


Fig. 7. Attentive response induced by robot behaviors. (A) Response by enthusiasm level; (B) Response by gender; and (C) Interaction effect.

gesture (included in all behaviors, $Mean = 4.32$, $SD = 0.98$), $utterance$ (included in behaviors with *moderate* and *high* enthusiasm, $Mean = 3.89$, $SD = 1.02$) and $approach$ movement (included in behaviors with *high* enthusiasm, $Mean = 3.60$, $SD = 0.55$) suggest that *Pepper*'s behavioral elements were recognizable. It is interesting to note that some participants who encountered the *mild* behavior indicated that they heard the robot speaking ($Mean = 3.44$, $SD = 1.13$). They claimed that their answers were based on previous encounters with *Pepper* at this location. It appears that they expressed what *Pepper* was expected to do (wave and speak) and not what it was really doing (wave).

Around 8.5% of the participants (Table I) displayed behaviors related to a novelty effect (Section V-A.2.a). The novelty factor decreased over time (Fig. 6) with a significant difference between the three phases of our experiment (one-way ANOVA, $F(2, 29) = 18.26$, $p < .001$). A post hoc Tukey test showed that phase 1 differed significantly from phase 2 ($p < .01$) and phase 3 ($p < .001$), with no significant difference between phases 2 and 3 ($p > .05$). These findings

TABLE II

MEASUREMENTS FOR THE PASSERSBY WHO ENTERED ALONE ($N$=230)

| Enthusiasm Level | Participants | | | Duration | | Walking Speed | | 5-Keypoint Score | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Total | Males | Females | Mean | SD | Mean | SD | Males | Females | Combined |
| Mild | 82 | 55 (67.1%) | 27 (32.9%) | 3.86 | 0.98 | 1.40 | 0.43 | 0.90 | 0.70 | 0.84 |
| Moderate | 73 | 53 (72.6%) | 20 (27.4%) | 3.73 | 0.86 | 1.42 | 0.37 | 0.77 | 0.75 | 0.77 |
| High | 75 | 65 (86.7%) | 10 (13.3%) | 3.98 | 1.20 | 1.39 | 0.47 | 0.95 | 0.90 | 0.95 |
| Combined | 230 | 173 (75.2%) | 57 (24.8%) | 3.86 | 1.02 | 1.40 | 0.43 | 0.88 | 0.75 | 0.85 |

support *H1* and show that the novelty effect exists. The tests further indicate that novelty was minimal during phase 3.

The passersby' reactive responses induced by the robot behaviors (summarized in Table I) represent the passersby' attentiveness (Section V-A.2.c). A two-way ANOVA test (with enthusiasm and gender as factors) showed that the main effect of enthusiasm (Fig. 7-A) was significant ($F(2, 358) = 4.98$, $p < .01$), whereas the gender effect (Fig. 7-B) was not ($F(1, 358) = 0.3$, ns). These findings imply that robot behaviors with *high* level of enthusiasm lead to a more attentive response from passersby, which supports *H2*. In addition, the interaction effect (Fig. 7-C) was found significant ($F(2, 358) = 3.1$, $p < .05$) and more pronounced for the female than for the male participants. A post hoc Tukey test showed a significant difference between *mild* and *high* enthusiastic behaviors for females ($p < .01$) but not for males ($p > .05$).

To further analyze the automated data collected by our system, we first extracted the records of passersby who entered alone (those who entered in groups of two or more will be presented in future work). Then, we removed 14 records for having errors related to getting the frames from the robot camera (i.e., where no or only a couple of frames were recorded). We also excluded the 18 participants who approached *Pepper* or stopped in front of it. The distribution of the remaining 230 records is summarized in Table II.

Using the collected head keypoints, we first analyzed the 5-keypoint score (Table II) as described in Section V-A.2.c. A two-way ANOVA test showed that the main effects of both enthusiasm ($F(2, 224) = 5.01$, $p < .01$) and gender ($F(1, 224) = 4.12$, $p < .05$) were significant (Fig. 8-AB). However, the interaction effect (Fig. 8-C) was non significant ($F(2, 224) = 1.25$, $p > .05$). Next, we analyzed the 5-keypoint time ratio (Section V-A.2.c). A two-way ANOVA test showed that the main effect of enthusiasm was significant ($F(2, 224) = 6.05$, $p < .01$). However, the gender effect was non significant ($F(1, 224) = 0.13$, $ns$), as well as the interaction effect ($F(2, 224) = 1.32$, $p > .05$). These findings support *H2* by showing that behaviors with *high* enthusiasm lead to more attentive responses. The 5-keypoint score also indicate that males and females reacted differently (with males expressing more attention).

When analyzing the walking speed of the passersby (Table II), a two-way ANOVA test showed that the main effect of enthusiasm was non significant ($F(2, 224) = 0.12$, $ns$). However, the gender effect was found significant ($F(1, 224) = 7.91$, $p < .01$), which showed a difference

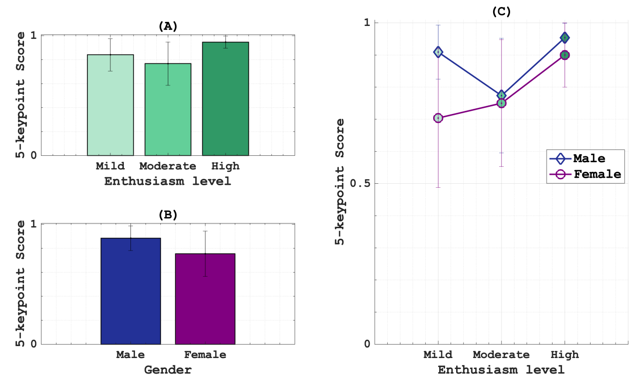


Fig. 8. 5-Keypoint score for the passersby. (A) By Enthusiasm level; (B) By gender; and (C) Interaction effect.

between the walking speed of males ($Mean = 1.36m/s$, $SD = 0.40$) and females ($Mean = 1.53m/s$, $SD = 0.46$). As for the interaction effect, it was found non significant ($F(2, 224) = 0.05$, $ns$). These results could be an indication that males were more attentive to the robot than females.

## VII. DISCUSSION

The analysis of the results in Section VI indicate that, in our setting, the novelty effect wore off over time and was minimal during phase 3 of our experiment. In addition, the passersby were able to recognize the role of the humanoid *Pepper* as a welcoming robot and appreciated its presence at the building entrance.

As for the passersby' attentive responses, the results indicate the presence of a significant pattern where attention was the highest when the behavior enthusiasm was *high*. It was also interesting to find a significant interaction effect for attentiveness between gender and behavior enthusiasm. It seems, therefore, that female and male subjects behave differently towards a welcoming robot. This finding was further corroborated by the significant effect of gender on walking speed. Further study is needed. In addition, deeper investigation is needed by using the findings from HHI studies (e.g., [30], [31], [14]) in order to develop more effective behaviors and improve engagement in HRI.

The limitations of this study are mainly related to the presence of robots in public spaces which is still novel and not very common. In addition, the behaviors we selected to measure the novelty factor (e.g., approach the robot or stop to watch) are relevant to our context and to the role of our robot (i.e., a welcoming robot). This may not be applicable for a different setting (i.e., where approaching the robot is needed). Finally, the current technology (i.e., vision techniques) - that we used to extract sensitive features from the passersby (e.g., head-keypoints) - is still limited and its performance depends on the context (e.g., lighting conditions and distance). The output from these techniques should hence be carefully considered when attempting to predict behavioral cues from a distance (e.g., gazing direction).

## VIII. Conclusion

In this study, we modeled a three-step model for drawing the attention of passersby and engaging them in the wild (i.e., outside the laboratory). This model was tested in a specific setting which makes its generalization somewhat limited (e.g., it may fail in a busy scenario). It can also be considered as the first step toward designing an attention-based engagement model for social robots. Furthermore, the effect of robot behaviors (with varying enthusiasm) for drawing the attention of passersby was investigated. The findings imply that robot behaviors with a *high* level of enthusiasm draw more attention than those with a *mild* or *moderate* level of enthusiasm. They also indicate that gender of participants influenced this effect. However, further studies are needed to validate these findings.

In a welcoming situation, the three-step model proved to be useful and effective for selecting and engaging visitors entering a building. This can be complemented by a follow-up strategy for approaching interested visitors based on the attentive response (e.g., walking behavior [24]). Future work includes extending this model to monitor the willingness to engage, before and after the robot's engagement behavior. This will be useful for selecting targets to engage as well as for maintaining the engagement or disengaging accordingly.

## References


[1] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro, "A two-month field trial in an elementary school for long-term human-robot interaction," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 962–971, Oct 2007.

[2] H.-L. Cao, P. G. Esteban, A. De Beir, R. Simut, G. v. d. Perre, D. Lefeber, and B. Vanderborght, "A survey on behavior control architectures for social robots in healthcare interventions," *International Journal of Humanoid Robotics*, vol. 14, no. 04, p. 1750021, 2017.

[3] C.-M. Huang, T. Iio, S. Satake, and T. Kanda, "Modeling and controlling friendliness for an interactive museum robot," in *Robotics: Science and Systems (RSS)*, Jul 2014.

[4] I. Aaltonen, A. Arvola, P. Heikkilä, and H. Lammi, "Hello Pepper, may I tickle you?: Children's and adults' responses to an entertainment robot at a shopping mall," in *Proceedings of the Companion of the ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 53–54.

[5] G. Castellano, A. Pereira, I. Leite, A. Paiva, and P. W. McOwan, "Detecting user engagement with a robot companion using task and social interaction-based features," in *Proceedings of the International Conference on Multimodal Interfaces*, ser. ICMI-MLMI. New York, NY, USA: ACM, 2009, pp. 119–126.

[6] D. G. Macharet and D. A. Florencio, "Learning how to increase the chance of human-robot engagement," in *IEEE/RSJ International Conf. on Intelligent Robots and Systems*, Nov 2013, pp. 2173–2179.

[7] M. P. Michalowski, S. Sabanovic, and R. Simmons, "A spatial model of engagement for a social robot," in *9th IEEE International Workshop on Advanced Motion Control*, 2006, pp. 762–767.

[8] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to approach humans?: Strategies for social robots to initiate interaction," in *ACM/IEEE International Conference on Human Robot Interaction*. NY, USA: ACM, 2009, pp. 109–116.

[9] C. Breazeal, *Designing Sociable Robots*. Cambridge, MA, USA: MIT Press, 2002.

[10] C. Shi, M. Shimada, T. Kanda, H. Ishiguro, and N. Hagita, "Spatial formation model for initiating conversation," in *Proceedings of Robotics: Science and Systems*, Los Angeles, CA, USA, June 2011.

[11] M. Finke, K. L. Koay, K. Dautenhahn, C. L. Nehaniv, M. L. Walters, and J. Saunders, "Hey, I'm over here - how can a robot attract people's attention?" in *IEEE International Workshop on Robot and Human Interactive Communication (ROMAN)*, Aug 2005, pp. 7–12.

[12] R. Gehle, K. Pitsch, T. Dankert, and S. Wrede, "How to open an interaction between robot and museum visitor?: Strategies to establish a focused encounter in HRI," in *ACM/IEEE International Conference on Human-Robot Interaction*. NY, USA: ACM, 2017, pp. 187–195.

[13] L. J. Corrigan, C. Peters, D. Küster, and G. Castellano, *Engagement Perception and Generation for Social Robots and Virtual Agents*. Cham: Springer International Publishing, 2016, pp. 29–51.

[14] A. Kendon, *Conducting Interaction: Patterns of Behavior in Focused Encounters*. New York, NY: Cambridge University Press, 1990, no. 6, ch. A description of some human greetings, pp. 153–207.

[15] P. Liu, D. F. Glas, T. Kanda, and H. Ishiguro, "Data-driven HRI: Learning social behaviors by example from Human-Human interaction," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 988–1008, Aug 2016.

[16] M. Balconi, *Intentions and Communication: Cognitive Strategies, Metacognition and Social Cognition*. Milano: Springer Milan, 2010, pp. 159–175.

[17] T. W. Liew, N. A. Mat Zin, and N. Sahari, "Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment," *Human-centric Computing and Information Sciences*, vol. 7, no. 1, p. 9, 2017.

[18] E. Saad, M. A. Neerincx, and K. V. Hindriks, "Welcoming robot behaviors for drawing attention," in *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2019, pp. 636–637.

[19] E. T. Hall, *The Hidden Dimension*. Garden City, N.Y.: Doubleday, 1966, ch. X: Distances in Man, pp. 113–129.

[20] J. Sung, H. I. Christensen, and R. E. Grinter, "Robots in the wild: Understanding long-term use," in *ACM/IEEE International Conference on Human-Robot Interaction*, March 2009, pp. 45–52.

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

[22] T. Kanda, M. Shiomi, L. Perrin, T. Nomura, H. Ishiguro, and N. Hagita, "Analysis of people trajectories with ubiquitous sensors in a science museum," in *IEEE International Conference on Robotics and Automation*, April 2007, pp. 4846–4853.

[23] S. Goldin-Meadow and M. W. Alibali, "Gesture's role in speaking, learning, and creating language," *Annual Review of Psychology*, vol. 64, pp. 257–283, 2012.

[24] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "A robot that approaches pedestrians," *IEEE Transactions on Robotics*, vol. 29, no. 2, pp. 508–524, April 2013.

[25] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human-robot interaction," in *IEEE International Conf. on Robotics and Automation*, vol. 4, May 2002, pp. 4138–4142.

[26] E. Torta, J. van Heumen, R. H. Cuijpers, and J. F. Juola, "How can a robot attract the attention of its human partner? a comparative study over different modalities for attracting attention," in *Social Robotics*. Berlin, Heidelberg: Springer, 2012, pp. 288–297.

[27] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2016, pp. 779–788.

[28] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *IEEE Conference on Computer Vision and Pattern Recognition*, July 2017, pp. 1302–1310.

[29] G. Trovato, J. Galeazzi, E. Torta, J. R. C. Ham, and R. H. Cuijpers, "Study on adaptation of robot communication strategies in changing situations," in *Social Robotics*. Springer, 2015, pp. 654–663.

[30] E. Goffman, *Behavior in Public Places: Notes on the Social Organization of Gatherings*. New York, NY: The Free Press, 1966.

[31] F. Poyatos, "Language and nonverbal behavior in the structure of social conversation," *Language Sciences*, vol. 4, no. 2, pp. 155 – 185, 1982.