The role of expressiveness and attention in human-robot interaction — Source link ⧉

A. Bruce, Illah Nourbakhsh, Reid Simmons

Institutions: Carnegie Mellon University

Related papers:

- A survey of socially interactive robots

- Emotion and sociable humanoid robots

- Designing Sociable Robots

- The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places

- Social interactions in HRI: the robot view

The Role of Expressiveness and Attention in Human-Robot Interaction

Allison Bruce, Illah Nourbakhsh, Reid Simmons

Abstract— **This paper presents the results of an experiment in human-robot social interaction. Its purpose was to measure the impact of certain features and behaviors on people's willingness to engage in a short interaction with a robot. The behaviors tested were the ability to convey expression with a humanoid face and the ability to indicate attention by turning towards the person that the robot is addressing. We hypothesized that these features were minimal requirements for effective social interaction between a human and a robot. We will discuss the results of the experiment and their implications for the design of socially interactive robots.**

Keywords— **human-robot interaction**

I. Introduction

THIS research is situated within a larger project with the ultimate goal of developing a robot that exhibits comprehensible behavior and is entertaining to interact with. Most robots today can interact only with their creators or with a small group of specially trained individuals. If we are ever to achieve the use of robots as helpmates in common, everyday activities, this restricted audience must expand. We will need robots that people who are not programmers can communicate with. Much work is being done on the side of receiving input from humans (gesture and speech recognition, etc), but relatively little has been done on how a robot should present information and give feedback to its user. Robots need a transparent interface that regular people can interpret.

We hypothesize that face-to-face interaction is the best model for that interface. People are incredibly skilled at interpreting the behavior of other humans. We want to leverage people's ability to recognize the subtleties of expression as a mechanism for feedback. This expression is conveyed through many channels: speech, facial expression, gesture, and pose. We want to take advantage of as many of these modalities as possible in order to make our communication richer and more effective. We also hope to discover in a principled way which ones are most significant and useful for human-robot interaction.

Most day-to-day human behavior is highly predictable, because it conforms to social norms that keep things running smoothly. When robots do not behave according to those norms (for example, when they move down a hallway swerving around human "obstacles" rather than keeping to the right and passing appropriately), it is unpleasant and unnerving. In order to be useful in society, robots will need to behave in ways that are socially correct, not just near optimality within some formal framework.

Following the line of reasoning above, it would be easy to say, "if making a robot more human-like makes it easier to understand, then the best thing to do would be to make an artificial human". Clearly this is not feasible, even if it were the right approach. But it does raise some useful questions. How anthropomorphic should a robot be? Can it be a disadvantage to look "too human"? If we can only support a few human-like behaviors, which are the most important for the robot to exhibit?

II. Related Work

There has been a significant amount of work towards making software agents that are believable characters who exhibit social competence. The projects such as the Oz Project [1] and Virtual Theater [8] created software agents that exhibit emotion during their interactions with each other and with human users with the goal of creating rich, interactive experiences within a narrative context. REA [4] and Steve [9] are humanoid characters that use multimodal communication that mimics the body language and nonverbal cues that people use in face-to-face conversations. While this work shares our goal of expressive interaction with humans, the characters are situated within their own "virtual" space, which forces people to come to a computer in order to interact. We are interested in developing characters that are physically embodied, capable of moving around in the world and finding people to interact with rather than waiting for people to come to them.

Work of this nature with robots is less developed than similar work with software agents, but it is becoming more common. There have been several museum tour guide robots designed recently to interact with people for educational and entertainment purposes. Nourbakhsh and collaborators at Mobot, Inc. address many of the same issues in human-robot interaction that we do in their discussion of their design de-

Carnegie Mellon University Robotics Institute, abruce@ri.cmu.edu, illah@ri.cmu.edu, reids@ri.cmu.edu

cisions, along with offering suggestions based on their experiences with several robots [13]. However, their primary focus was on using entertaining interaction to support their educational goals rather than conducting an in-depth study of face-to-face social interaction. Minerva, another museum social robot, used reinforcement learning to learn how to attract people to interact with it, using a reward proportional to the proximity and density of people around it [12]. The actions that the robot could employ for this task included head motions, facial expressions, and speech acts. Their experimental results did not show that certain actions were more successful than others with any statistical significance other than that friendly expressions were more successful at attracting people than unfriendly ones.

Kismet is a robot whose sole purpose is face-to-face social interaction [3]. It uses facial expressions and vocalizations to indicate its emotions and guide people's interaction with it. Kismet is specifically designed to be childlike, engaging people in the types of exchanges that occur between an infant and its caregiver. In contrast, our goal is to engage people in a dialog similar to an interaction between peers, using expressiveness to support our communicative goals. Another major difference between this project and ours is that Kismet is a head and neck on a fixed base. Even though Kismet is a physical artifact, like the software agents mentioned above, it relies on people coming to it in order to engage in interaction. While our robot is stationary for this particular experiment, one of the goals of this project is to explore the effects of an agent's ability to move around freely on the quality of social interaction with it.

III. System

Our testbed is a RWI B21 equipped with a laser range finder. A pan-tilt device with a flat screen monitor attached is mounted on top of the robot. The screen is used to display the robot's face, which is an animated 3D model. Speech and the accompanying phonemes ,which are used for lip-syncing, are generated by the Festival [2] text-to-speech software package. The use of a rendered face allows us more degrees of freedom for generating expressions than would be possible if we designed a face in hardware.

The face design that we are currently using for our robot, Vikia, is that of a young woman. This initial design was chosen because we hypothesized that a realistic humanoid face would be easier for people to interpret the expressions of, and we wanted the robot to appear non-threatening. Later we hope to try a number of different facial designs and compare their relative merits.

The facial expressions that Vikia exhibits are based on Delsarte's code of facial expressions. Francois Delsarte was a 19th century French dramatist who attempted to codify the facial expressions and body movements that actors should perform to suggest emotional states [10]. He exhaustively sketched out physical instructions for actors on what actions to perform, ranging from posture and gesture to fine details such as head position and the degree to which one should raise their eyebrows to indicate emotion. His approach, designed for melodramatic stage acting, is well suited for our application because it is highly systematic and focused on the communication of emotional cues to an audience. We focused our attention on the portion of Delsarte's work that dealt with facial expressions and head position. An animator implemented facial expressions for many of the more common emotions (happiness, sadness, anger, pride, shame) that Delsarte codified on the model for Vikia's face. For each emotion, Delsarte's drawings indicate the deformations that must be made to the facial features to express that emotion at varying levels of intensity. We created facial expressions for Vikia at 3 intensity levels for each emotion we implemented. These facial expressions are used to add emotional displays to Vikia's speech acts. The robot's speech and the animation of the head and face are controlled using a scripting language that allows for the sequencing of head movements, facial expressions, and accompanying speech. The language represents behaviors as state machines that transition on signals sent by the programs that manager perception. This allows new robot behaviors to be developed with relative ease. The script for the experiment was created using this system.

Vikia is equipped with a laser range finder, which we use to track the location of nearby people. The tracker runs at 8 Hz and is capable of tracking an arbitrary number of people within a specified area (set to a 14 ft radius around the robot for the purposes of this experiment). Occlusion often makes reliable detection of every person walking together in a group impossible. The tracker will always succeed in detecting a group of people as the presence of at least one person, however, which is adequate for the performance of this task.

IV. Experiment

The task that the robot performed was that of asking a poll question. There were a number of reasons for choosing that task. From an implementation point of view, it is a short and very constrained interaction, so it can be scripted by hand relatively easily. And the feedback that the robot needs to give in order to appear that it has understood the human's response is minimal (a necessity for now, as we have not yet

integrated speech recognition into our system). Also, because people are egocentric and interested in sharing their opinions, we believe that we can expect a reasonable degree of cooperation from participants. Taking a poll contains many of the elements of interaction we are interested in studying (particularly the aspect of engaging people in interaction) without having to deal with the complexity of a full two-way conversation. We think that success at this task will indicate a significant first step towards longer, more complicated, and more natural interactions.

The robot's script for the poll-taking task ran as follows. First, the robot waits to detect that someone is in its area of interest. When the robot detects someone, it greets them and begins tracking them. The robot will pay attention exclusively to this person until the interaction is finished. If the person stops, the robot will ask them if they will answer a poll question. If they are still there, the robot will ask the poll question, asking them to step up to the microphone (mounted on the pan/tilt head) to answer. If the person does not step forward, they will be prompted to do so 3 times. If the person hasn't cooperated by then, the robot tells the person that it is giving up on them and ends the interaction. Once the person steps forward, the robot detects that they are within a threshold distance, which the robot interprets as a response to the question. Because there is currently no speech recognition onboard the robot, this is the only available cue that the person has answered. The robot waits for the person to step back outside of this threshold. If they fail to do so, they are prompted them to step back 3 times before the robot gives up. Once the person is outside the threshold, the robot determines that the interaction is over, thanks the person, and says goodbye. The interaction is then repeated with the next nearest individual.

We observed the number of people that passed by, that the robot greeted, that stopped, that responded to the poll question, and that finished the interaction. The response variable recorded for this experiment was whether or not a person stopped when greeted by the robot. This number provides a measure of success at attracting people to interact, rather than of the success at completing the interaction. Relatively few people out of the number that stopped actually completed the interaction. The two major reasons for this were that people could not understand the robot's synthesized speech and that people did not step in close to the robot to answer, so the robot would prompt them to step closer. They would answer more loudly from the same distance and become frustrated that the robot could not hear them.

A. Experiment Design

We were interested in exploring the effects of the expression of emotion and indication of attention on the robot's success at initiating interaction. Without the face or the ability to move, the robot relies solely on verbal cues to attempt to engage people in interaction. Passersby receive no feedback on whether the robot is directly addressing them if there is more than one person walking by at a given time. By turning towards the person it is talking to, the robot removes this ambiguity. Also, gaze is an important way that people initiate interaction with others, so this cue should be recognizable and familiar to people. The face offers an additional level of expressiveness through the accompaniment of the speech acts by facial expressions (the output of the speech synthesis package that we use is not modulated to indicate emotion) and supports people's desire to anthropomorphize the robot. Would people find interaction with a robot that had a human face more appealing than a robot with no face? Previous work on software agents suggests so [6] [11], even indicating that people are more willing to cooperate with agents that have human faces [5].

The emotions that the robot exhibited during this interaction were all based on its success at accomplishing the task of leading a person through the interaction. Vikia greeted passersby in a friendly way. If they stopped, Vikia asked the poll question in a manner that indicated good-natured interest. If the person answered, Vikia stayed happy. But if the person didn't behave appropriately according to the script (for example, if they didn't come closer to answer or stayed too close and crowded the robot) Vikia's words and facial expressions would indicate increasing levels of irritation. This proved to be fairly effective in making people comply or attempt to comply with Vikia's requests. However, people who didn't step closer to answer and spoke louder instead often seemed perplexed and offended by the robot's annoyance with them.

The experimental design was that of a 2x2 full factorial experiment, a common experimental design used to determine whether the factors (variables) chosen produce results with statistically significant means and whether there is an interaction between the effects of any of the factors [7]. The factors that we manipulated were the presence the face and having the robot's pan/tilt head track the person's movements.

The robot was placed in a busy corridor in a building on the CMU campus. We acknowledge that CMU students, particularly most of the ones that are in the computer science buildings, are not a representative sample of the general population. Our rationale for choosing to do the experiment on campus is that the sheer novelty of having a robot in a public place is usu-

ally enough to attract most people. At CMU, seeing robots is more typical, so people will be less likely to stop to interact overall. But it is important to note that this shouldn't have an effect on people's reaction to the factors that we are interested in testing.

A.1 Factors

• Face. The robot's face in this experiment was an animated computer model of the face of a young woman displayed on a flat screen monitor mounted on the pan-tilt head of the robot. When the face was not used, the screen was turned off.
• Tracking. The robot uses a laser range finder to locate and track the position of a person's legs. Using this information, the robot can turn the screen towards the person that it is interacting with and follow their motion.

A.2 Schedule

This experiment was conducted over a period of four days with 2 trials in the morning and two in the afternoon. Over the course of the experiment, each combination of factors was tested in each trial time as well as on each day. We included factors for the time of day and the day of the trial during our analysis of the data in order to determine if effects due to time had an impact on our experiment.

V. RESULTS

The results obtained for the effect of each factor individually are shown in figure 1. The dependent variable is expressed as a person's probability of stopping (calculated from the experiment data).
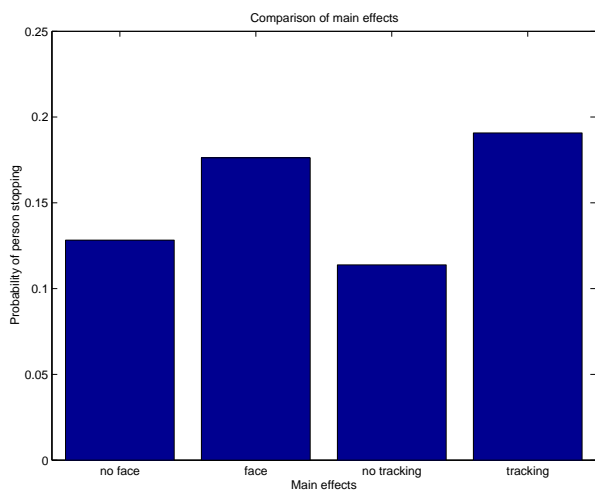

Fig. 1. Main effects of face and tracking.

The data was analyzed using analysis of variance (ANOVA) for all factors. In analysis of variance, F-

Source	P-Value	Confidence
Main effects		
Tracking	0.002	> 99%
Face	0.042	> 95%
Interactions		
Face x Day	0.014	> 95%

tests are performed in order to determine whether the differences between the mean values for the factors (or combinations of factors) are statistically significant. Our results indicate that both the face and the tracking behavior had statistically significant effects, with over 95% confidence (p = .042) for the face and over 99% confidence (p = .002) for tracking (see table 1).
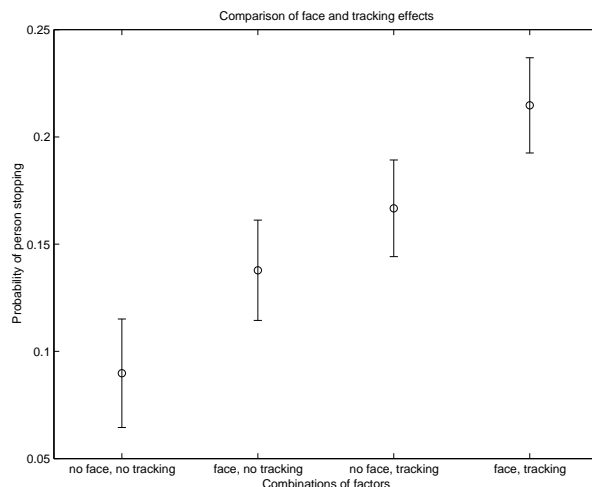

Fig. 2. Interaction between the face and tracking with standard error intervals.

The analysis of variance also revealed an interaction effect between the face and day factors. This means that these variables effected each other in a systematic way. In this case, the use of the face produced less of an effect on people's willingness to stop during trials conducted later in the course of the experiment as opposed than it did at the beginning. We hypothesize that this was due to some kind of habituation effect. While we assumed for the purposes of experimental analysis that our data was independent, in reality there was some repeat traffic through the hallway during the week the experiment was taking place. It seems that the face may have been less effective at getting a person to stop and interact a second time than the tracking behavior. While there isn't sufficient information to draw any conclusions about this effect, it

raises some interesting questions. Is this relationship particular to our experimental conditions, or does it reflect larger differences in the importance of physical movement versus anthropomorphism for social tasks?

The results indicate no interaction between the face and tracking (e.g., the difference between the percentage of people who stopped to interact with the robot when it had a face and when it did not was roughly the same regardless of whether the robot was tracking them, even if more people stopped overall when the robot was tracking them as well). This suggests that while both expression and attentive movement are important on their own, their combination results in the most compelling behavior, giving a roughly additive increase in performance (see figure 2).

VI. Future Work

This work is in its preliminary stages, and there are numerous promising directions we hope to explore. It is obvious that this kind of interaction would benefit from richer sensing, such as speech input and visual cues. Explicitly modeling common social behaviors, such as approach and avoidance, and using these models to reason about people's intentions could also vastly improve the quality of interaction. Additionally, we plan to test people's reaction to less passive forms of robot motion, such as the robot approaching people whom it is trying to interact with.

VII. Conclusions

We have performed an experiment on the effects of a specific form of expressiveness and attention on people's interest to engage in a social interaction with a mobile robot. The results of this initial experiment were both encouraging and surprising. They suggest that having an expressive face and indicating attention with movement both make a robot more compelling to interact with. Furthermore, the use of both together yields a roughly additive increase in performance at our experimental task. A number of questions were raised that have yet to be explored, both about our design and implementation and the assumptions that motivated it. In future work, we will continue to experimentally test our theories about what features and abilities best support human-robot interaction.

References

[1] J. Bates, *em The Role of Emotion in Believable Agents*, Communications of the ACM 37 (7), 1994, 122-125.

[2] A. Black, P. Taylor, and R. Caley, Festival Speech Synthesis System http://www.cstr.ed.ac.uk/projects/festival/festival.html

[3] C. Breazeal. and B. Scassellati, *How to Build Robots That Make Friends and Influence People*, In Proceedings of IROS-99, Kyonju, Korea.

[4] J. Cassell, T. Bickmore, H. Vilhjlmsson, and H. Yan, *More Than Just a Pretty Face: Affordances of Embodiment*, In Proceedings of 2000 International Conference on Intelligent User Interfaces, New Orleans, Louisiana.

[5] S. Keisler and L. Sproull, *"Social" Human Computer Interaction, Human Values and the Design of Computer Technology*, B. Friedman, ed. CSLI Publications: Stanford, CA.: 1997, 191-199.

[6] T. Koda and P. Maes, *Agents With Faces: The Effect of Personification*, In Proceedings of the 5th IEEE International Workshop on Robot and Human Communication(ROMAN 96), 189-194.

[7] Irwin P. Levin, *Relating Statistics and Experiment Design*, Thousand Oaks, California. Sage Publications: 1999.

[8] B. Hayes-Roth and D. Rousseau, D. *A Social-Psychological Model for Synthetic Actors*, In Proceedings of the Second International Conference on Autonomous Agents, 1998, 165-172.

[9] J. Rickel, J. Gratch, R. Hill, S. Marsella, and W. Swartout, *Steve Goes to Bosnia: Towards a New Generation of Virtual Humans for Interactive Experiences*, In papers from the 2001 AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment, Technical Report FS-00-04. Stanford University, CA.

[10] G. Stebbins, *Delsarte system of dramatic expression*, E. S. Werner: New York, 1886.

[11] A. Takeuchi, and T. Naito, *Situated Facial Displays: Towards Social Interaction*, Human Factors in Computing Systems: CHI'95 Conference Proceedings, ACM Press: New York, 1995, 450-455.

[12] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. B.Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, *Probabilistic Algorithms and the Interactive Museum Tour-Guide Robot Minerva*, International Journal of Robotics Research 19(11), 2000, 972–999.

[13] T. Willeke, C. Kunz, and I. Nourbakhsh, *The History of the Mobot Museum Robot Series: An Evolutionary Study*, In Proceedings of FLAIRS 2001, Key West, Florida.