Robot Design Rubrics for Social Gesture Categorization and User Studies with Children

Heather Knight, Angela Chang, Walter Dan Stiehl, Robert Toscano, Yi Wang and Cynthia Breazeal
Personal Robots Group
MIT Media Lab 20 Ames St. Cambridge, MA 02142
{rehtaeh, anjchang, wdstiehl, rtoscano, yiw, cynthiab}@media.mit.edu

ABSTRACT

In order to improve integration and acceptance of robots in everyday society, this paper discusses a rubric to socialize robot design. Our innovative approach views robot design as an inclusive and revealing process. In the course of creating a real-time social touch architecture for a robotic teddy bear, we conducted an observational pilot study to understand how children interact with robots. We observed use of social tactile gestures to formulate and support our taxonomy of social touch for the teddy bear.

In this paper, we reflect on the reasons for the positive user reactions to this pilot system based on the features of the study scenario. We present some key elements of our experimental design which target non-traditional user study atmospheres that mirror the robot's target social context. We demonstrate that our method can be used to iterate subunits of robot design within a larger social context. We believe these design principles may also apply to various realms of social robot integration and support sustained improvement and development.

Categories and Subject Descriptors

[**HRI Communication**]: Conveying Intention, Gaze and Gestures, [**Robot Perception & Prediction**]: Modeling social situations.

General Terms

Measurement, Performance, Design, Human Factors, Verification.

Keywords

Robotic touch, Social Robot Design, Human-robot interaction, Metrics for Social Acceptance, Social Gesture Categorization, Gesture library, Robots for/with children, Robotic Companions.

1. INTRODUCTION

We are at the dawn of the age of Personal Robots, where everyday robots will become common in our homes and form part of our workplace and lives. This new class of machines will need to go beyond task performance to understanding human social interfaces and gestures[1]. This paper addresses two human robot interaction challenges; (1) How to meet the need for introducing robots into real life, and (2) How the development of robotic agents can be sustainable.

Figure 1.a) Sensate Bear b)Siblings interact with during Study

This research forms part of the Huggable project, which is a personal robot platform in a teddy bear form factor for healthcare, education, and entertainment ([2],[3]). In prior work with the Huggable, we have demonstrated a diverse set of multi touch interactions on a single paw segment using off-line techniques ([4][5]). This work demonstrates the next step, a real-time recognition of full-body touch based on behavioral and user studies with adults and children.

We created a parallel research plan to create a social touch taxonomy. The specific aim of this investigation was to characterize nonverbal touch communication toward a robotic teddy bear. Our studies evaluated the nature, role, and utility of social touch[1] between humans and robotic companions ([6]).

The Sensate Bear system (see Figure 1a) uses a network of 56 capacitive sensors based upon [2]. Due to space limitations, the reader is encouraged to see our companion paper [7], which describes the technical design in more detail. This system is used to sense and characterize tactile interactions by the use of sensors distributed over the full-body teddy bear form.

We present first observational studies using the Sensate Bear test rig with children (Figure 1b). The data was used to first formulate and then validate our social touch taxonomy. Our iterative process allows social interaction to be incorporated into a system's design at the earliest implementation. It also demonstrates participatory user acceptance during development of a larger robotic system.

2. BACKGROUND

The first robots were industrial robots, and that has shaped much of our approach to robot design. Even in the realm of human-robot interaction, researchers tend to describe challenges in the context of team collaborations. Our scheme begins to address a new way of thinking about a robot as a participatory member, proposing the idea of the robot as a family-friendly entity. We are interested in how this entity's presence can changes relationships and interactions between the people in the room.

Traditional HRI studies are one-on-one human-robot interactions inside a sterile lab environment. There have been some studies involving dyadic interaction within a home [8], but these

---

[1] We define *social touch* as touch that performs a social function, such as hugging for comfort or poking to get attention [6].

environments have been task-based. One important aspect of our approach is to recognize how prior works may assume paradigms of successful performance exist and underemphasize sociability.

The home environment for which personal robots are designed should extend beyond dyadic interaction. Tanaka et al put a robot in an immersive classroom environment ([9][10]). Our work extends their investigation to acknowledge that a whole family unit will interact with the robot. Thus, in our study we actively invited children and their family members to be present in the room to interact with the Sensate Bear.

3. METHODOLOGY

Our project acts as an iterative case study for design based on observed human actions. We initially examined how adults used gestures to develop a socially structured taxonomy of touch. This taxonomy was used to design processing algorithms that were consistent with our observations and measurements[7]. We later validated the taxonomy through observational user studies with children.

3.1 Initial Observational Study

Before hardware construction, we observed users in simulated target situation in order to determine sensor placement and processing based on behavioral study. We used an ordinary plush teddy bear to characterize an initial set of tactile interactions (Figure 2). This bear was the same model which was ultimately used as the exterior covering for the Sensate Bear. This early study was conducted with adults, 5 female and 4 male volunteers.
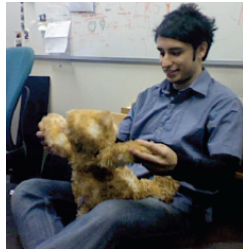

Figure 2. Adult participant with plush bear

The experimenter facilitated a role-play scenario between the bear and eight participants, describing a scenario, e.g., how the bear is feeling and activities the plush bear wanted (e.g., read a book or play with an object). The participant acted out the interaction and the experimenter occasionally projected the plush bear's verbal reaction to the participant actions during the study. Participant behaviors were video recorded, with interactions lasting about five minutes. The clips were later tagged to identify touch gestures, contextual intentions, related regions and significant/insignificant areas. The processing procedure is depicted in Figure 3.
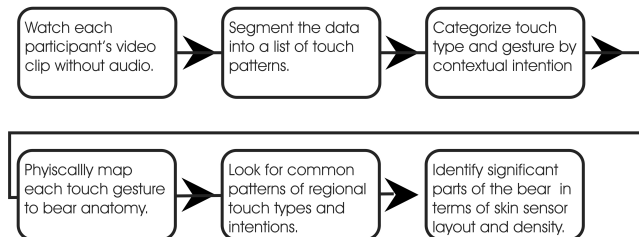

Figure 3. Procedure used for creating touch taxonomy from study data

All participants completed the full set of situational exercises. The bear was in physical contact with all participants at least 95% of the time, though there was no prompting in that regard. Six out of the seven treated the bear like a social creature throughout the experiment, positioning and manipulating the bear in orientations standard for babies and living creatures. The only exception being one participant that tried to tie the bear into a knot. All made eye contact with and talked to the bear.

3.2 Social Touch Taxonomy

Developing a socially-structured vocabulary for robot social touch is necessary to better integrate robots into human social environments. For example, if a robot did something correctly, it would need to be able to understand that an head pat is affectionate (good work) rather than attentional (look here). As regards types of touch, the initial observational study displayed the gesture categories shown in Table 1.

Table 1. Five observed gesture categories for touch interaction

Affection	Manipulate	Puppeteer	Attention	Playful
head patting, hugging	moving, positioning, supporting	act out response with bear	gesture or poke for attention	tickling, scratch back

Focusing on gestures with social content, we found both *symbolic gestures* that have social significance across individuals and associated regional touch distributions (e.g. hug, foot rub), and *touch subgestures*, which are smaller scale and are independent of location (e.g. pat, poke). The most commonly observed symbolic gestures were hugs, head-pats, tickle, shake awake, attentional tapping, petting, hand shake, and holding hands. Subgestures included tickle, poke, pet, and hold. For example, the tickling subgesture often occurred in the underarms of the bear, but the bear could also be tickled by touching other regions (e.g. foot).

3.3 Contextual Design of Sensor System

Our initial study observations informed the sensor parameters. Relevant region sizes ranged from approximately 2x2" in a head pat, to three simultaneous regions in a hug, see Figure 3. Touch duration and migration were on the order of 2-5 seconds and 5 inches respectively. The highest density touch-locations were the sides and underarms, the top of the head and the shoulders and arms. Regions touched less often in this study included the back, feet and face of the bear.

The sensor layout was mapped to an anthropomorphic physical layout, so that discrete bear regions mapped to body parts (e.g. head, arm, etc.). This anthropomorphic organization gave a useful mapping for programming, as the gestures could leverage human experience of these gestures (e.g. a hug involves front and back regions, rather than thinking of sensors 1-20). Each bear region contained a subset of sensors, with a dedicated processor for that region. A main sensor board processed data from activated regions to classify gestures. A visualization is shown in Figure 4.

Thus recognition of local gestures, or *touch subtype*, and *social gestures* were detected and statistically classified in real-time. The features used to distinguish pet, pat, tickle and hold include amplitude, frequency spectrum, base frequency and duration of the signal at each sensor. We trade 'perfect characterization' for real-time classification (80% accurate). In our application, characterized gestures and activated regions were rendered on a 2D graphic of the bear in real-time.
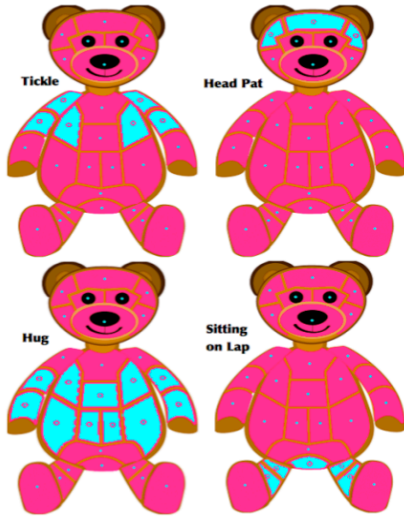
Figure 4. Visualization of Symbolic Gesture Filters for Tickle, Headpat, Hug and Sitting on Lap

3.4 User Studies for Hardware Evaluation

In order to verify the ability of the hardware to classify gestures in real-time, the bear was used in a preliminary study involving children. We figured children could demonstrate basic social gestures and local subtypes. We used the bear without fur, and tracked un-amplified sensor signals while recording video.

We invited eleven children to participate in a largely qualitative study with the foam Sensate Bear. All were under the age of twelve. The participants were aged 4-11, seven boys and four girls. None hesitated to engage with the bear and their touch was similar to how they would treat a real creature, involving all surfaces of the bear. We also invited the family (parents and siblings) to participate in the room with the child. This method brings users into the design process and better captures unexpected emergent data. The family members often encouraged a shy young sibling to touch the bear through demonstration and showing their own interest in the project.

3.4.1 Interactivity Modes and Procedure

We explored three techniques for evaluating the interaction between robot and users. In the *performance mode*, the robot was audio-puppeteered. In *exhibition mode*, users were exposed to the inner workings of the bear through a realtime on-screen visualization of active sensors and gesture categorizations. A third *mixed mode*, used both the audio-puppeteering and visualization.

Study variables included: presence of audio-puppeteering, visibility of sensor activations, and duration of interaction. We also tracked the number of children in the room, gender and age. Evaluation variables that were recorded included: eye contact (bear, study conductor, visualization, parents), initiation of new touch gestures, position of bear (lap, table, in arms), verbal interview of reactions. Upon the child's arrival the study conductor would follow the steps in Figure 5. The study conductor was always in the room with the child, as were the parents and any siblings. The audio puppeteer was not in the room. The study duration was under ten minutes.
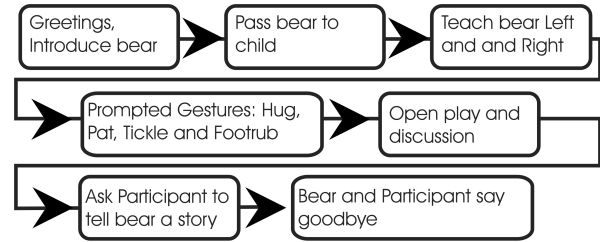

Figure 5. Steps used in preliminary user study

3.4.2 Gesture Categorization Results

The most common bear positions observed (from most to least common) were: Sitting on lap, Sitting on table in front of child, Held in arms, Lying on lap, Lying on table, Held in air over table. Common bear manipulations: Pick-up, Sit back on table, Make bear dance. Table 2 lists the performance of the sensed gestures.

Tickle, Head-pat, Foot-rub, and Hug were the specific classifications evaluated. We chose these because they were the most commonly occurring kinds of symbolic touch as identified in the initial behavioral study. Additional observed gestures included handshake, belly tickle, foot tickle, shake awake, go to sleep, lay to sit, feeding and rocking.

Table 2. Initial Gesture Recognition, Accurate Activations

Gesture	First try	w/Explanation	Region accuracy
Headpat	100%	100%	100%
Tickle	20%	60%	20%
Hug	40%	80%	80%
Footrub	100%	100%	100%

Head-pat and Foot-rub were consistently identified by the system for all users. Hug was sometimes identified, but using the unamplified signal meant that the sensors did not detect contact with clothing, a problem less relevant in the next hardware evaluation as the video data appears to map to classification locations. Contrary to expectation, Tickle did not have a consistent locational distribution, with subjects 'tickling' feet, stomach and neck in addition to the more common adult mapping of tickle to the underarm and side regions.

3.4.3 Response to Interactivity Modes

Older children were most engaged in the social touch gestures, or *performance mode* of the bear while the youngest were most fascinated with the on-off reactions of the visualization in *exhibition mode*. All responded enthusiastically in both cases, eliciting a variety of distinct interactivity content.

Performance Mode: All children played along with the audio-puppeteering and after the first minute treated the bear as if it were speaking. Procedural conversation (exchanging names, information) provoked fewer touch interactions than laughing or sound effects (e.g. bear snoring). These tendencies may indicate that (1) conversations of a more emotional nature, particularly those involving the emotions of the bear are well associated with a symbolic touch response (reassurance, affection) and (2) conversations involving basic living functions (sleeping, eating, tummy-ache, laughing) provoke more of a caretaker response (rocking, feeding, stroking).

Exhibition Mode: In the visualization case, subjects were consistently engaged in testing the functionality and reactions of the bear. When given the visualization, they learned those skills quickly and seemed to enjoy the visual response, particularly

among the younger audience. When it was off, the gestures were not activated as reliably. The action-reaction testing became a game to the children, and their discovery of a gesture, as mentioned above, provoked excitement and further curiosity, as they sought to retrigger the gesture label and discover even more.

Mixed Mode: One or the other sensory interaction dominated during mixed mode. Either the child became less interested in the conversation as they looked more at the screen than the bear, or less interested in the visuals as they almost forgot the physical presence of the bear in the flow of the conversation. This varied with the age of the child and the content of the task.

3.5 Reflections on User Studies with Children

In the course of conducting this study, we began to identify unique elements that were useful in our user study with children.

Family Involvement: Children do not like to be alone. Including parents and siblings helps them open up faster, introducing a group dynamic. The older the child (there were three subjects over the age of ten), the faster they understood the project and began interacting. In contrast, those under five waited until prompted to touch the bear. The family was an encouraging support for those initially shy children. This atmosphere fostered social relationships and incorporated surprises. There was an element of fun and comfort during the experiment where users did not feel like specimens but seemed more like contributors to the project.

Multisensory interaction: Further, multisensory interaction made the process engaging for children who might have short attentions. The use of voice created a sense of continuity and story context for the children. The real-time visualization made it enjoyable on a quick-impulse time scale. They were very comfortable with it (and older siblings, in particular, even tried to foil/hack the bear to see if they could fool its detection).

Open-ended context: The talking/chatting nature of the study, rather than task performance made the study relevant and useful for the entire 4-11 year old age group. Using a form factor that was not threatening helped us avoid the uncanny valley. The familiar teddy bear shape elicited real enthusiasm from children, and drew on their experience with prior teddy bear toys.

Seeing Under the Hood: Revealing the workings of the bear (via real-time visualization) opened up the level of conversation about the technology. The technical nature of robotic design usually excludes the user, but we found that people treated the experience more like an exploratory museum exhibit. They were interested in the research done in this part of robots functionality and design, a small subset of a much larger system.

4. EVALUATION AND CONCLUSION

Our study was not an attempt to completely characterize the new rules for social robot user studies and user studies with children. However, we do propose a vision for how to conduct studies that move beyond the traditional machine-user dyad and performance metrics. We encourage other researchers to push forward with this idea and expand on some of these methods in their own work.

This approach is different because (1) it uses the family context, (2) it opens the hood and lets the users see how the robot works on the inside, and (3) allows for flexibility in how the users interpreted the experience. We believe this participatory context was rewarding beyond any tangible benefit they might have

received as study subjects. The response of the families was enthusiastic, and we expect that the experience will make them more accepting to personal robots in the future. Through educating the participants about ongoing robotics research, we believe our work addresses new rubrics for the sustainable development of sociable robotic agents.

Our approach is to include users in the design process as early and often as possible. From behavioral observation to participatory evaluation, we found that the intended end-users enjoyed being part of the design process. We have demonstrated creative methods for introducing interactivity, iterating design, and revealing the workings of research prototypes as a way to introduce more social contexts in the development of personal robot systems.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] McGee, K., "A touch of the future: contact-expressive devices," *Multimedia, IEEE* ,11 (1) , Jan.-Mar. 2004 , 8-11.

[2] Stiehl, W.D. et al. 2005. Design of a Therapeutic Robotic Companion for Relational, Affective Touch. IEEE Int'l Workshop on Robot and Human Interactive Communication (RO-MAN 2005), 408-415. Field, T. 2001. Touch. MIT Press.

[3] Toscano, R. L, Lee, J.K., Stiehl,W.D .; Breazeal, C. 2008. The design of a semi-autonomous robot avatar for family communication and education. Robot and Human Interactive Communication, 166-173.

[4] Stiehl, W.D., Breazeal, C. 2006. A Sensitive Skin for Robotic Companions featuring Temperature, Force, and Electric Field Sensors. Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2006), 1952-1959.

[5] Stiehl, W.D., C. Breazeal, "Affective Touch for Robotic Companions," in First International Conference on Affective Computing and Intelligent Interaction, Beijing, China, 2005.

[6] Knight, H. 2008. An Architecture for Sensate Robots: Real Time Social-Gesture Recognition using a Full Body Array of Touch Sensors. Master's Thesis, Dept. of EECS. MIT.

[7] Knight, H., C. Breazeal. Technology and Algorithms for Real time Touch Gesture Recognition on a Robot with a Full Body Sensate Skin. IROS 2009. Currently under review.

[8] Kidd, C., C. Breazeal. A Robotic Weight Loss Coach. Twenty-Second Conference on Artificial Intelligence. Vancouver, British Columbia, Canada. AAAI, 2007.

[9] Tanaka F., Cicourel, A. & Movellan, J. R. 2007. Socialization Between Toddlers and Robots at an Early Childhood Education Center. Proceedings of the National Academy of Science. Vol. 194, No 46.

[10] Tanaka F, et al. Daily HRI evaluation at a classroom environment: Reports from dance interaction experiments. In Proceedings of the 1st Annual Conference on Human-Robot Interaction, pages 3–9, 2006.