

*An Architecture for Sensate Robots:
Real Time Social-Gesture Recognition using a
Full Body Array of Touch Sensors*

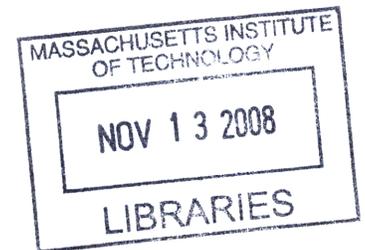
Heather-Marie Callanan Knight
B.S., M.I.T., 2006

Submitted to the Department of Electrical Engineering and Computer Science in Partial
Fulfillment of the Requirements for the Degree of

Master of Engineering in
Electrical Engineering and Computer Science at the
Massachusetts Institute of Technology

August, 2008

©2008 Massachusetts Institute of Technology
All rights reserved.



Author _____
Department of Electrical Engineering and Computer Science
August 31, 2008

Certified by _____
Cynthia Breazeal
M.I.T. Thesis Supervisor

Accepted by _____
Arthur C. Smith
Professor of Electrical Engineering
Chairman, Department Committee on Graduate Theses

Outline

I. Abstract

II. Background

III. Applications

III. The Sensate Bear: Implementation and Studies

A. Construction: A Robot with Full Body Sensate Touch

B. Salient Touch Gestures: A Behavioral Study

C. Touch Gesture Recognition Design

D. Software Implementation for Real-Time Recognition

E. Utility and Acceptance: A Pilot Study with Children

IV. Discussion and Conclusions

V. Future Work

VI. Acknowledgements

VII. References

Appendix A. Behavioral Study Script

Appendix B. Sensate Bear Study Fliers

I. Abstract

Here we present *An Architecture for Sensate Robots: Real Time Social-Gesture Recognition using a Full Body Array of Touch Sensors*, as part of Heather Knight's Masters of Engineering in Electrical Engineering thesis work in August 2008 with Cynthia Breazeal as thesis advisor.

Touch plays a central role in social expression but, so far, research into social touch behaviors for robots has been almost non-existent. Embodied machines have the unique capability to sense human body language, which will enable robots to better comprehend, anticipate and respond to their human companions in a natural way.

This thesis addresses the novel field of sensate touch by (1) creating the first robot with full body sensate touch and with on-screen visualization, (2) establishing a library of salient social gestures through behavioral studies, (3) implementing a first-pass touch gesture recognition system in real-time, and (4) running a small pilot study with children to evaluate classifications and test the device's acceptance/utility with humans. Such research is critical path to conceiving and advancing the use of machine touch to better integrate robots into human social environments.

All of the above will be incorporated into the Huggable robotic teddybear at

the MIT Media Lab's Personal Robotics group and makes use of the *Sensitive Skins* circuit design created in Dan Stiehl's Masters thesis [1]. This implementation substantially reduces his proposed total sensor numbers and type, modularizes sensors into two uniform shapes, and extends his valuable work on a single body section to an evaluation of sensors over the entire surface of the robot.

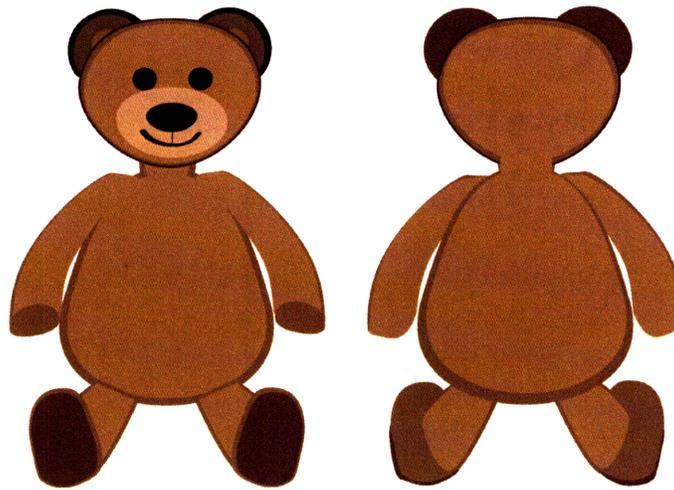


Figure 1: Visualization Bear

The sensate bear presented here consists of a network of 58 capacitive sensors that detect the presence and proximity of humans by measuring our effect on the sensor electronic field. These sensors are particularly well suited to social gestures, because they will not sense most inanimate objects such as chairs or blankets or walls due to their non-conductivity.

Humans are conductive due to our high water content, thus we act as a second electrode plate to the sensor electrode, changing the effective

capacitance of the overall system, an effect that is translated into a changing signal

voltage, which a microcontroller converts into a number readable by computer program. By individually querying an array of sensors distributed over the surface bear at a high cycle rate, the robot attains a high accuracy dynamic model of where and how



Fig. 2: Sensate Bear

it is being touched at a particular instant of time.

A real-time sensitive skin system for robots would allow both researchers and robot to characterize touch's role in social and emotional communication and thereby improve human-robot interaction. The goal of this project is to create a 'Version 1.0' Touch Module and API implementation that is robust and expandable and that can be easily integrated into the Huggable behavior system. With that, one can begin to explore the impact of touch on engagement level and social communication by characterizing and contrasting the sociability of a robot with and without sensate touch.

What distinguishes this implementation and research is that the sensors are

full body and that the research question it emphasizes involves characterizing full body social gestures , such as head pats, tickling, hugs, in addition to the detection of modes of affective touch. The basis of these gestures came from an initial behavioral study and its final puppeteered interactions with children provided good news for robot touch, as the children's affection and openness to the bear gave valuable information about future training.

II. Background

This section describes the previous work done on social machines in the Personal Robotics group of the MIT Media Lab and introduces the Huggable Project. Next, it highlights the rising importance of machine-human social interfaces, then surveys the role that touch traditionally plays in human to human social communication, and in the final segment, reviews related research using robotic touch.

The Huggable and its Predecessors

The touch implementation here is intended for the Huggable robot, a project at the Personal Robots Group in the MIT Media Lab. The group is headed by Professor Cynthia Breazeal, who pioneered the field of sociable machines with her landmark PhD project Kismet, who responded to and

incited human emotion, and book *Designing Sociable Robots*, in which she defined a generalized architecture for creating sociable machines [2]. Using similar design principles, the touch-based analysis in this thesis should feel like a natural extension of the way we would expect a teddy-bear to interpret and respond to our gestures.

In addition to the short-term interactions, one of the latest innovations to come out of Personal Robots is the idea of tracking the long-term status of the human-robot relationship. First conceived by Cory Kidd, this new variable involves changing the behavior of the robot based on how long the robot and person have known each other and how well they are 'getting

along.' It was inspired out of an extension of human psychology research for Autom, his weight management robot, when he was thinking about what kinds of social cues users would find



Figure 3: The Huggable

natural and helpful when using the system daily over a months to years timeframe [3]. Touch has long been one of the clearest measures of social comfort with another person, and could provide a great additional angle in furthering this research.

The most important robot for affective touch, however, is the Huggable (pictured right), the robotic teddybear which is the center of this thesis. It was originally conceived by Walter Dan Stiehl of the Personal Robots group around 2005 and is intended for use in a range of healthcare, remote-learning and remote-communication applications.

The 2006 IEEE Consumer Communications and Networking Conference included the following abstract for the Huggable, "Numerous studies have shown the positive benefits of companion animal therapy. Unfortunately, companion animals are not always available. The Huggable is a new type of robotic companion being designed specifically for such cases. It features a full body sensitive skin for relational affective touch, silent, muscle-like, voice coil actuators, an embedded PC with data collection and networking capabilities." [4].

Since then, the group has also begun collaborations and research exploring its potential for use as a Robot Communication Avatar and as a Early Education Companion. All of these will be predicated on the successful implementation of touch based behaviors, based on the next generation skin boards, which were installed over the entire surface of a robot for the first time (previous research was limited to a single arm segment). Other sensors likely to be useful in behavior characterization are joint potentiometers and the Inertial Measurement Unit, which is also an active research module on

the bear.

The Rising Importance of Human-Robot Social Interfaces

As we enter an era where human-robot collaboration and day-to-day interaction gains prevalence, it becomes important for machines to learn to navigate social interfaces. Rather than teach people how to talk to and relate to machines it will be much more effective to have machines that can understand humans on human terms. As such, we can teach, direct and involve robots in our daily rhythms in the same way that we would pets, children or each other without extensive training or unnatural interfaces (like a mouse and keyboard). If machines could better understand and interpret social cues, their ability to learn tasks and help us accomplish our goals would accelerate significantly and society could better and more easily adapt to their inclusion.

What are the benefits of better communication and understanding between humans and machines?

- Better task definition and learning
- More propensity to look to a machine as a partner in accomplishing a task
- Friendlier, more-effective working relationships with machines

The extensive benefits of an embodied machine to help us with physical tasks and operate in the 3D world is in great part what distinguishes a robot from its paraplegic computer predecessors. So far, most development in human-robot communication interfaces has focused on interpreting verbal commands and vision processing, which is a natural outcropping of technology's previously immobile animations. On a television and computer screen only the audio-visual was possible. With robots, touch and movement become natural therefore assumed capabilities and we should take full advantage of the new possibilities this creates for machine-human communication.

New interfaces call for new innovations and a brief reflection on how we relate to our pets, children and friends quickly brings to the forefront the centrality of touch to social and emotional interaction as well as our acknowledgment and attachment to physical objects.

We are evolutionarily wired to interpret systems that give evidence of social behaviors as social creatures. Humans are wonderful pattern detectors, we can see paths in the woods that evade virtually all computer vision systems and we can single out single voices and conversations at a crowded cocktail party. In the same way, from a young age and with only a few social cues (take teddy bears as evidence) we find it easy to make emotional bonds to objects.

In a recent study conducted by Cory Kidd with Autom, a robotic weight-management coach with a humanoid face and face-tracking eyes, there were participants that named their robot or even bought Redsox gear for them to wear. Of the 45-person 6-week-long study, fifteen participants received robots with touchscreens to help them track their fitness and nutrition patterns, fifteen just touchscreens and fifteen kept pen and paper journals. There was no evidence of similar social relationships developing between the participants and weight-management tools in the second two test groups [3].

Fellow researcher, Guy Hoffman became very aware of this phenomenon when he created the control architecture for the robot-actor AUR's debut in

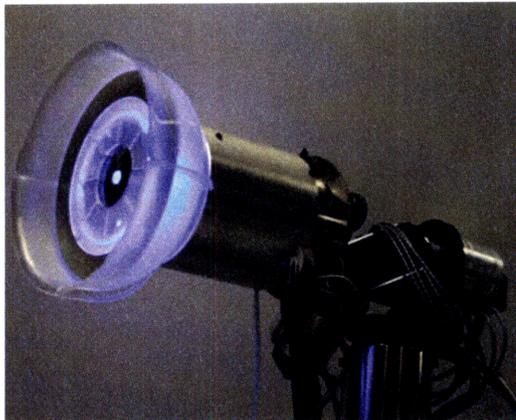


Figure 4: AUR

a play called *The Confessor* this past spring at MIT (pictured left).

He noted that one of the first things the actors wanted to do when they were relating to AUR, particularly during intense moments of the performance was to touch the

ne, ignoring the potential for erratic motion as they became immersed in AUR's character's responses, though only expressed through color and simple movements [5].

Humans are fundamentally social creatures. We evolved in an environment where relating to each other and a speedy understanding of the subtleties of emotion and social minutiae were critical to our

immediate survival. Language is a key component in how we express and exchange ideas but it is also rather inefficient and seldom captures the full meaning of what we are trying to communicate. Identical word sequences can translate to a full spectrum of distinct meanings, including polar opposites in the case of sarcasm, if there is a change of context, tone of voice or different body language accompanying the sentence. Furthermore, if an animal is about to pounce on a friend, pulling them out of the way or a quick shout, point, and distressed expression are going to be much more effective at saving them than calmly and clearly explaining the threatening nature of the situation.

Rather than tackle all the non-verbal subcomponents of human social interaction at once, this thesis seeks to expound upon a novel machine-human social interface that has only become possible with the rise of the embodied computer, the machine that can not only move but sense.

Philosophers have long explored the question of the mind/body distinction and what bonds the individual to society at large. As individuals, our window to the outside world is filtered by our ability to sense it - the cones and rods in eyes color our surroundings, cochlea hair cells help distinguish

vibrations in the air as musical pitch and, of course, a range of cells in our skin perceive and interpret different kinds of touch. As Marshall McLuhan declared in his 1964 opus on the emerging mass-media, “the medium is the message” [6] and thus, it is the sensory channels we give machines to explore its environment that will define and limit its abilities to operate in and understand its world in the same way we do.

The Role of Touch in Social Communication

Touch is central to human interaction and the way we relate to both objects and each other, but it is a sense whose power we often underestimate. For example, one study showed that waitresses who unobtrusively touch their customers receive larger tips than those who do not, even when customers rate the quality of the service at the same level afterwards and often do not consciously remember that they were touched [7]. According to Dr. Shelby Taylor, Psychology Professor at California State University, this is unsurprising as touch and charisma are closely intertwined, and she has created a training program based on her study of the body language of charismatic people [8].

A few examples of everyday sociable touch might include the following:

- Your dog wants to go outside and play so he puts his head on your

lap and nudges your arm away from the mouse but you brush him away.

- You show up at a salsa dancing class, the music blasts and you try to follow the pressure of the instructors arms and match the steps as he swirls you around the floor.
- Your niece won't pronounce her first full sentence until her third birthday, but she leads you by the hand to her favorite exhibit at the Museum of Science, invariably crying when she has to leave unless you pick her up and promise to return soon.

In all of these cases, animals, adults and children use touch to mediate social communication, whether it be to request, teach, express emotion or console.

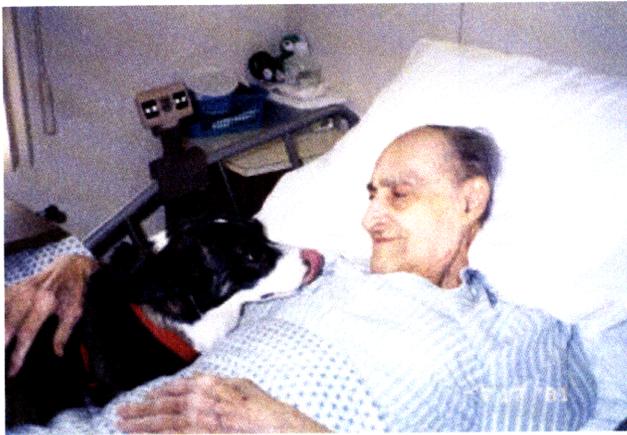
In addition to touch's usefulness in communication, the benefits of passive touch to a person's psychology and health has been experimentally tested. Dr. Tiffany Field's is the Director of the Touch Research Institute at the University of Miami and she has found that proper doses of touch and massage show benefits in attention difficulties, psychiatric illness, autoimmune disorders and pain by increasing serotonin production and decreasing stress hormones. Furthermore, the contrast of infants raised with foster parents to though subject to those in an orphanage, where there is less touch and individual attention shows higher IQ as well as faster weight gain and growth [9].

Thus, being touched (passive touch) in an appropriate context has positive benefits, but what about active touch, when one is touching others?

Evaluating Pet Therapy research gives much evidence to the benefits to active touch. Studies show that elderly people with pets often have lower stress levels, are less lonely and have a greater sense of companionship than

those who do not have pets.

Similarly, many hospitals now include pet therapy programs where animals are brought in to visit patients.



It appears that the simple act

of petting and eliciting an

Figure 5: Pet Therapy

emotional response from the

pet has a soothing impact on the patient, in addition nurturing a social bond

in what can often be a lonely medical environment [10]. As Dr Fred

Goodwin, Professor of Psychiatry and Director of the Center on

Neuroscience at The George Washington University Medical Center relates,

touch has traditionally been part of the helping professions, from doctors

making diagnosis to teachers encouraging students with a hug, although it is

becoming more rare in the rushed modern world and with increasing use of

technology-based tests in diagnosis [11].

Touch Research with Robots

In the world of engineering and robotics, the concept of incorporating touch has generally been limited to specialized designs for precision or safety and is only just beginning to be explored as a medium for sociability. In fact, the field of sociable robots and affective computing is also in its infancy. The general rule to date is that machine touch has not been 'affective,' and affective technologies have seldom used touch. There are, of course, a few exceptions.

One simple but intriguing use of touch in a social robot was in a five-month study led by Javier Movellan at the University of California, San Diego, in which a giggling



Figure 6: Giggling QRIO

QRIO robot was eventually treated as a peer in a classroom of preschoolers.

One of the unique characteristics of the robot was that it had a tactile sensor on its head and would start giggling when it was touched. According to National Academy of Sciences paper [11] about the project, "The results highlighted the particularly important role that haptic behaviors played in the socialization process: (i) The introduction of a simple touch-based

contingency had a breakthrough effect in the development of social behaviors toward the robot. (ii) As the study progressed, the distribution of touch behaviors toward the robot converged to the distribution of touch behaviors toward other peers. (iii) Touch, when integrated over a few minutes, was a surprisingly good predictor of the ongoing quality of social interaction."

This passage highlights the power of touch to both enable and provide indication of social engagement. The giggling response in conjunction with ambient roming behaviors, occasionally sitting down, and lying down when its batteries ran out that provided the full impression of creature-ness. In contrast to many previous studies, the children did not become less engaged with the robot over the course of the long-term study.

Further evidence of how much QRIOs programming changed the toddlers' perception of the robot's sociability was reported in the New Scientist, "The children also treated QRIO with more care and attention than a similar-looking but inanimate robot that the researchers called Robby, which acted as a control in the experiment. Once they had grown accustomed to QRIO, they hugged it much more than Robby, who also received far more rough treatment... Eventually, the children seemed to care about the robot's well being. They helped it up when it fell, and played "care-taking" games with it most commonly, when QRIO's batteries ran out of juice and it lay down, a

toddler would come up and cover it with a blanket and say "night, night"
[12]. So simple an implementation would be unlikely to have the same hold
over an adult, but a more complex system, such as the Huggable, that
incorporates sociable touch along with its other features is moving in the
right direction.

A related implementation explicitly created for affective touch research is
the Hapticat, an expressive wizarded device that could purr, change ear
position and exhibit different breathing rates depending on the interaction
mode of the study-participant, as observed by a human on another side of a
partition [13]. The participants were asked to exhibit ten different actions,
including pet, tickle, poke, hug, and each behavior would elicit a specific
response from the Hapticat.

According to the researchers, "It was particularly interesting to watch their
reactions the first time the Hapticat began to respond to their actions. Nearly
all exhibited strong positive reactions. One participant began to laugh so
hard that tears came to her eyes, and she was unable to report her responses
until she took a short break to regain her composure. The vast majority of
participants remained genuinely excited and engaged with the Hapticat
during the length of the study." They found that their behavior mappings,
especially for pet-owners, met user expectations for the responses they
would expect from the Hapticat and furthermore, that the participants

emotional responses were heightened when the Hapicat's actions matched those expectations.

The Hapticat was a low-complexity implementation with somewhat arbitrary response mappings, but what it lacks in autonomy is precisely the innovation that this thesis implements, namely, a full body sensate skin that can characterize and qualify different classes of touch without a human in the loop.

In total, we see that even in simple implementations, artificial touch-based affective devices are capable of engaging its users and successfully communicating an emotional response.

III. Applications

There are many areas where a sensate bear be useful or beneficial. Examples follow in hospitals, social disorders, education and long distance communication. The Huggable project has begun development of several of these already only lacking the full-body sensate skin.

Patient Diagnosis

Help diagnose patient condition by asking patient to show a doctor on the bear where and how they feel or hurt. By tracking the location and intensity of the touch one can have a physical understanding of the patients state that is more precise and useful than simply a verbal description. This will be useful in both traditional hospitals and psychiatry.

Social Learning and Behavioral Therapy

Reinforce positive social behavior and treatments by encouraging respectful interactions with the bear and discouraging hurting or misusing the bear.

Teach children positive social habits early such that they can better integrate with their peers. Can also be specifically targeted at children with social disorders, such as autism, who would particularly benefit from the instruction and attention of an intelligent social agent.

Distance Communication

Give feedback to a person remotely operating the bear about the communication partner's local tactile response. We are already developing techniques to remotely puppeteer the Huggable, but how should the user puppeteering the bear receive feedback about the interaction?

Take the example of an aunt talking to her young niece. If the aunt was remotely puppeteering the bear to communicate with the child, why not have her wear a jacket that translates the nieces touch gestures into vibration patterns on the garment. She could thus receive indications of hugs, attentional taps or affective states of her niece, thus closing the loop of affective communication. This is particularly useful when interacting with very young children, whose mastery of verbal communication has not yet developed.

Hospitals: The Child-Robot-Nurse Triad

The bear may better understand the child's emotional or physiological state than the nurse, either because it is in constant company with the child or because it is wirelessly away of patient monitoring devices. An autistic child may not show signs of anxiety on the outside, however, if the bear can sense the increased heartrate it can communicate that heartbeat to the nurse through an added vibro-tactile motor in its hand. Also, if it can sense a seizure or other serious medical condition, it can contact the nurse over the wireless network. Even playful situation where the robot can help explain a medical procedure coming up, or pretend to be taking a pill that the child does not want to take will aid hospital staff and help bridge the gap between the adult and child world.

III. The Sensate Bear: Implementation and Studies

The Huggable itself has already gone through several iterations and version 3.5 is currently under construction. Thus, these touch experiments were developed on a separate test-rig bear equipped exclusively with capacitive touch sensors, joint potentiometers, and a few modes of active response. The sensor-boards were created in two standard sizes and installed more or less uniformly over the bear surface for a total of 64 sensors.

The initial Huggable sensate skin design called for 1200 sensors of three different types, however, this bear simplifies the hardware and data processing design, allowing for the characterization of, 1) what kind of local touch is it (caress, poke, squeeze), 2) what the distribution of touch implies about the symbolic gesture being communicated (hug, tickle, head pat) and 3) what the affective content of that touch is (sad, angry, nervous, happy).

The identification of salient locations was based on a behavioral study where subjects role-played social interactions with a teddybear on videotape.

More chronologically, the general touch-behavior progression included:

1) Identifying natural and desired touch-based behaviors for the bear

Determining what those behaviors are came out of a combination of pet

therapy and mother-infant interaction research as well as videotapes of users interacting with the bear. That background will help determine the most salient areas on the bear, in other words, it will help determine which regions should have the highest sensor density and which regions may not require sensors for initial behavior categorization.

To get a quick cross-section of likely touch interactions with the bear, I had adult colleagues engage in a list of role-playing tasks with a standard teddybear. The interactions were videotaped and gestures tagged. This tagging helps quantify the ways people hold and interact with the bear. A more comprehensive analysis would compare and contrast similar tactile relations between parents and infants, people and pets. Instead, for simplicity, I will use documentation and everyday knowledge of how those interactions occur to further inspire and characterize the most important kinds of touch and help the bear better understand and react to the ways in which a human companion is likely to engage it.

2) Implementing the passive sensate test-bear, which will be used for data collection and first pass touch characterization

The test-bear will not move on its own but will have sensors distributed in a non-uniform network over its surface that can record its interaction with other people and will have the same size and shape as the original bear. The next version of the Huggable is also in development, thus the intention is to

have the skin module research progress independently until the later stages, when they will join together.

Figuring out where the sensor distribution is predicated by the saliency analysis in part one and intended to give an idea of overall touch interactions. The two major challenges here are the hardware itself and the the real-time classifier.

3) Running user studys to characterize touch gestures, interaction modes and verify system

There were two studies, one with adults and a standard teddy bear and the second with children and a sensate bear. In both cases the bear was passive, though in the second there was audio puppeteering and an on screen visualization. The former was used to determine initial sensor distribution and salient regions and the second to verify the utility and acceptance of the system with children. Both cases helped us better understand standard interaction modes with the bear under different bear conditions.

III-A. Construction: A Robot with Full Body Sensate Touch

This section describes of the current Sensate Bear hardware and software.

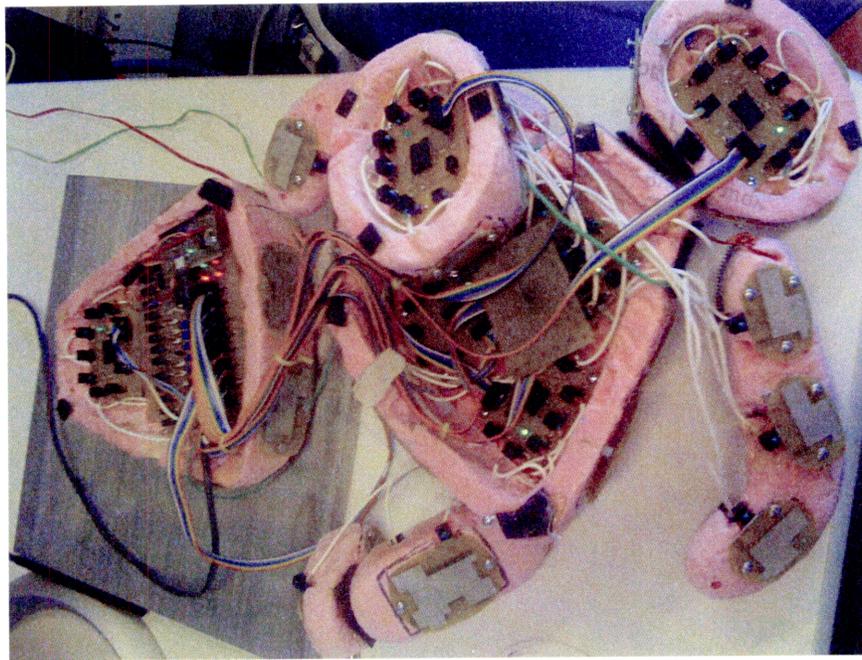


Figure 7: Sensate Bear Electronics

It is intended to be both a functional description of the first instance of a full body robotic skin developed for social touch, laying the groundwork for future development, and also as a user guide to the current system.

Functional Overview

Within the bear, all communication and calibration takes place on a centralized *Somatic Processing Board*. This board takes in data from seven *Midplane Boards* (there is capacity for eight), each of which is responsible for outputting the signal from up to eight capacitive sensors on the surface of the bear. Thus, there are currently 60 active *Capacitive Touch Sensors* distributed over the surface of the foam bear, of which 56 are used.

There are two supplies required to power the electronics, 12V for the capacitive sensing chip on each of the Midplane Boards and 5V for the microcontroller and signal conditioning / calibration circuitry on the Somatic Board. The microcontroller on the Somatic Board currently streams data from all sensors to a computer using serial over USB, thus connection is made with a USB cable.

Once on the computer, the Sensate Bear software, created with Microsoft Robotic studio in C#, reads the associated COM port data, performing thresholding, gesture recognition, and processing before displaying active sensors and gesture classification on a locally hosted website visualization.

Physical Structure

In order to ensure proper sensor activation, it is important to install the sensors close to the surface of the

bear. As a pilot study for the larger Huggable project, the necessary form factor was in the shape of the bear.

Thus, I constructed the foam such that it would fit under the fur of the commercial teddybear who was used in



Figure 8: In-Process Foam

the first three Huggable prototypes.

I also mirrored the Huggable mechanics to allow joints at each limb, foot and at the neck, the foam pieces include head, torso, right and left legs, arms and feet, as depicted in the diagram to the left. In order to accommodate electronics, the head and torso are hollow and made up of two pieces, front and back.

Thus, after initial measurements and designing the sensor layout, I used a hot-wire foam-cutter to rough out shape of each body section. I then used a large raster and finally sandpaper to create edges, curves and finer shapes. To hollow the inside, I used a half-inch drill-bit, also finishing the internal surface with a small raster and sandpaper.

The inter-segment connections have gone through a few iterations. The front and back head and torso pieces are connected with Velcro, as is the head to the body, though the head is additionally held in place by the wide ribbon cables connecting the head Midplanes to the centrally located Somatic. Ultimately, neck, limb and feet joints should be equipped with potentiometers or even servos in the case of the arms and head. So far, the limbs are connected with loops of metal sculpture and with loops of stiff fabric and Velcro allowing for more natural bending. The natural flexibility of the loop construction reduces friction between segments, allowing for

smooth motion, while constraining the limits of its rotation and the wire keeps the limbs from coming loose when the Velcro fails.

Power Supplies

The bear uses a twelve and five volt supply. The former is to power the

capacitive sensing chip and the latter to power the Somatic multiplexers, operational amplifiers, Serial-over-USB chip and PIC 16F877A microcontroller. The current implementation uses a power supply, but could easily operate off portable batteries.

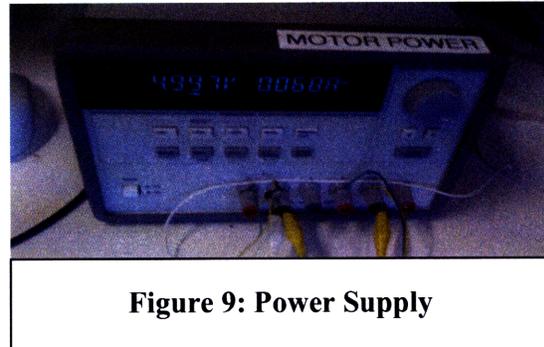


Figure 9: Power Supply

Power (P) consumption is calculated from the average currents (I) on each supply, which are 70mA on the 5V supply and about 30mA on the 12V. Thus, as $P=IV$, $P = 0.07\text{Amps} * 5\text{V} + 0.03\text{Amps} * 12\text{V} = 4.6 \text{ Watts}$.

Sensor Electronics: Electrodes and Midplanes

The electronics architecture of the Sensate Bear is an adaptation of the circuitry developed by Dan Stiehl of the Personal Robots group. I worked with undergrad Yi Wang to do the design, layout and construction of the boards themselves. I was responsible for the overall architecture, dimensions and much of the final construction, while her invaluable efforts

consisted Eagle layout, coordinating with the manufacturer. We tested calibration circuitry together with Dan Stiehl's counsel.

A diagram showing the main elements of the bear electronics connected together follows below. It consists of a tree structure with the Somatic Processing Board at the root. The Somatic has connectors for up to eight Midplane boards, which can each manage eight Sensor boards.

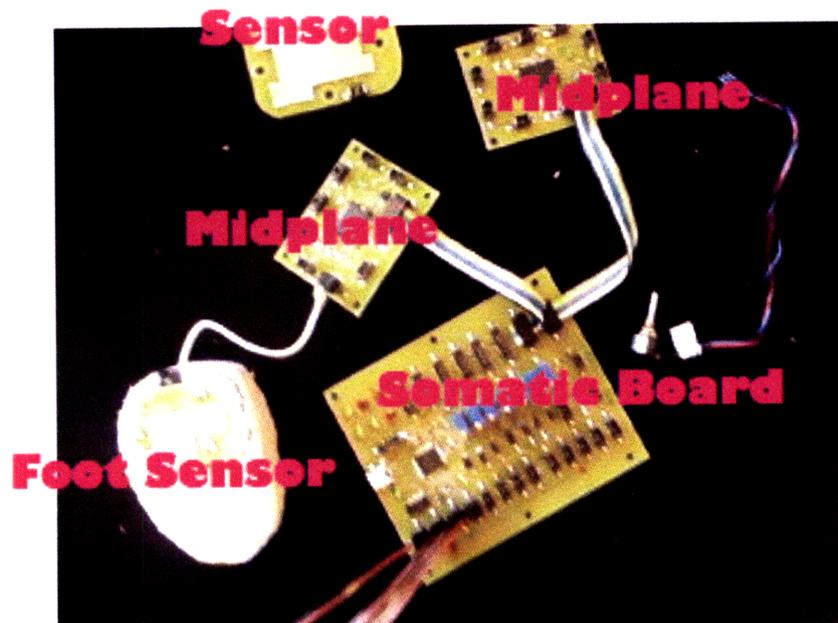


Figure 10: Sensor, Midplane and Somatic Boards

The sensor boards have electrode plates on both sides consisting of large pads of copper and a connector. One side is signal and faces outward, the other side is shield and faces inward. The shield helps direct the sensitivity of the sensor to the outside surface of the bear, amplifying the signal thereof as well as reducing the likelihood of sensor cross-triggering and interference from the electronics on the inside of the bear.

The Midplanes are dominated by a capacitive sensing IC with internal eight-channel muxing as well as a shield and sensing signals which set the oscillation of the respective sensor board electrodes. “Signal” and “shield” are always active for all channels but which channel is processed and transmitted to the midplane is set by a three bit control signal, indexed at one, from the Somatic. So that means 001 is channel 0, 010 is channel 1, 011 is channel 2 and so forth.

To prevent confusion, note that inactive (no touch) signals are high and active (touch) signals are low. The shield signal usually centers around 3.8V and the sensing signal usually centers around 4.2V.

Capacitive Sensing Circuitry Debugging

If the sensing signal is not coming through as expected, there are a set of tests that have almost always found the culprit. If any of the items below do not check out, continuity test connectors, correct mux signals and verify there are no shorts on the midplane or sensor board, particularly between and shield and electrode signal.

Midplane Signal Debugging Checklist:

1. 12V supply
2. Ground Connection
3. Correct Electrode Channel Selected
4. Electrode centered ~4V
5. Shield centered ~5V

The signal from the sensing electrode travels to the Miplane board and then to the Somatic Processing Board before amplification. To prevent interference and signal loss, it is important that the signals be processed as close to the sensor as possible. More details in Dan Stiehl's Sensate skin Thesis.

The Somatic Processing Board

The Somatic is responsible for communications and managing the signals

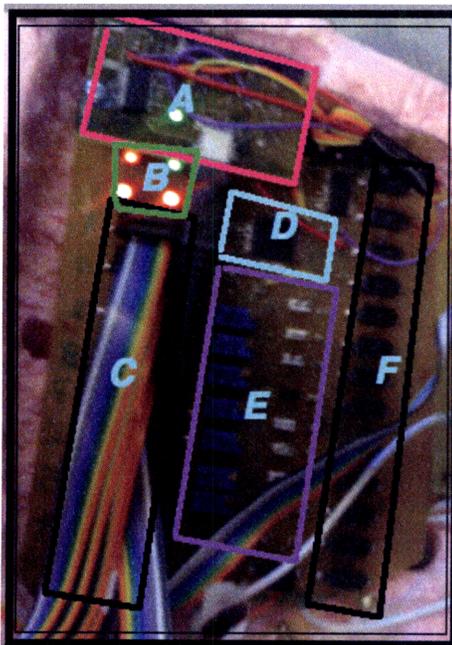


Figure 11: Somatic Subsections

and power for the whole bear. A diagram labeling the sub-sections on the somatic is shown below. **Section A** Contains the serial-over-USB circuitry, chip and mini-USB connector. **Section B** shows four of the six indicator LEDs. The three green indicator LEDs correspond to 5V, 12V and ~5V in, the last useful when using an unregulated supply, as it indicates the level going into the voltage regulator (unused currently). The three red

LEDs represent the three bits of the current midplane multiplexer signal.

Normally the flickering of all but the third will be too fast for the eye to see.

Section C contains the miplane connectors.

Section D is the PIC 16F877A microcontroller and related circuitry.

Section E contains the calibration circuitry (see next section).

Section F contains thirteen connectors for the joint potentiometers.

Currently unused but functional.

The opposite side of the board has the signal multiplexers and operational amplifiers as well as some of the additional passive components needed by the sections detailed above.

Signal Calibration on the Somatic

The original signal will only have a 5-30 millivolt fluctuation through the fur, so if we want to use the sensors with calibration circuitry, we must

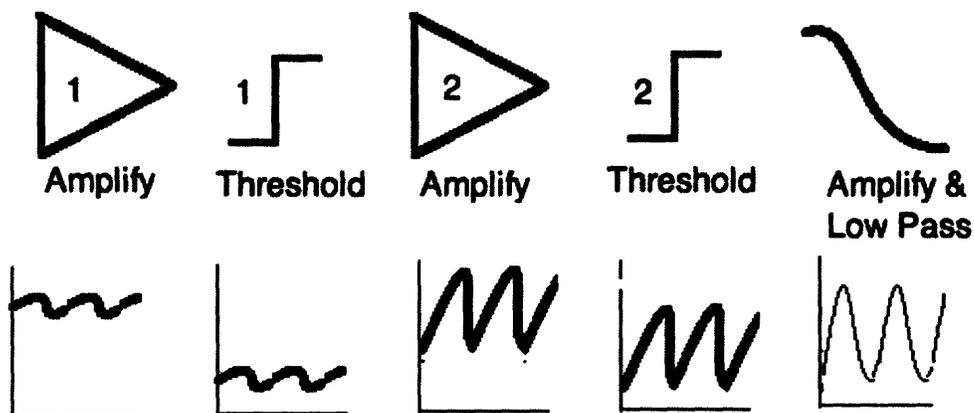


Figure 12: Signal Calibration

amplify and process the signal. The calibration circuitry implemented on this somatic consists of the following stages, all of which can be jumpered and reconfigured if only a subset is required or for debugging purposes.

The input comes from the midplane and the output goes to the touch sensor A/D input pin on the microcontroller.

Connector Construction



In search of more resilient connectors that could be rapidly unplugged and plugged in without concern for wires breaking (a departure from previous implementations), I elected to use Tico MTA-50 headers and connectors.

We used three wire connectors for the

Figure 13: Connector

potentiometers, two wire connectors for the power, seven wire connectors for the midplane to somatic, and modified two-wire connectors for the touch sensing.

After several trials, I believe to have found an almost unbreakable technique for ensuring everything stays together, by applying epoxy to the cable tension caps before pressfitting the ribbon cable wire onto the connector.

The technique became particularly tricky for the Capacitive Sensing single wires, because they need to be shielded with signal running down the inside and shield connected to the shield. Thus one must first connect tapered two-wire ribbon cable to the shielded cable, where the longer lead goes to ground and the shorter to the internal signal (this configuration reduces the likelihood of accidental shorting). After one does this on both sides, keeping the connector «legs» facing down in the direction of the wire (see diagram), one can proceed as normally with the final epoxy and pressfitting.

III-B. Salient Touch Gestures: A Behavioral Study

This study was designed to ballpark and begin characterization of tactile interactions that people have when playing with a teddybear, particularly, a teddybear that can interact with its users. It was intended to give an initial



Figure 14: Behavioral Study

estimate of important gestures rather than a rigorous analysis of all gestures, as behavior toward the bear is likely to vary as its designs and interactive capabilities evolve.

Experiments used a traditional teddybear in role-playing situations, with participants who were familiar with the idea of the Huggable and some of the goals of having a robotic teddybear. While some might question the idea of using participants with a knowledge of the subject, our participants' exposure to the concept of a robotic teddybear may have predisposed them to adapt more readily to having an engaged two-way interaction (at least in role-play) with teddybear.

It would also be beneficial to conduct a similar study with children to evaluate the potential differences and overlap between adult and children populations in this context, which was one of the motivations of the pilot described in section III-D.

Procedure of Behavioral Study

The bear used was the same bear model that was used for the fur of the original Huggable. Each participant interacted with the bear individually and on videotape and the clips were later evaluated to identify touch

gestures, contextual intentions, related regions and significant/insignificant areas. There were 7 sitting participants and 1 standing.

The protocol of the experiment was as follows:

- 1) Have participant sit or stand in place and begin videoing
- 2) Study conductor hands participant the bear, prompted to say hello
- 3) Conductor runs through a list, telling participant how the bear is feeling or what the bear wants to do, occasionally 'puppeteering' the bear's verbal reaction.
- 4) Conductor concludes experiment, asks participant for further suggestions about modes of interaction.

The worksheet used by the conductor for this experiment is in Appendix A.

The processing procedure was as follows:

- 1) Watch each participant video-clip individually with audio off
- 2) Segment data into a list of touch gestures
- 3) Categorize touch type by its contextual intention, this step is based on the coder's behavioral analysis, necessarily subjective
- 4) Map each touch-gesture to locations on bear, draw related region on sketch of bear with numerical label and category dependent color

- 5) Look for patterns of regional touch types and intentions
- 6) From above, infer both the most and least significant parts of the bear in terms of skin sensor layout and density

Behavioral Study Results

After observation, I classified contextual intentions into the following categories:

1. Affectionate Touch (head patting, hugging)
2. Manipulated Touch (moving, positioning or supporting the bear)
3. Puppeteering, in which the user puppets a socially relevant bear response
4. Attentional Touch (poking or slapping to get the bears attention, pointing bear arm)
5. Playful Touch (tickling, scratching back)

Gender#	Affectionate	Manipulated	Puppeteered	Attentional	Playful
Female 1	4 (31%)	4 (31%)	3 (23%)	0 (0%)	2 (13%)
Male 1	2 (22%)	2 (22%)	3 (33%)	1 (11%)	1 (11%)
Female 2	10 (42%)	5 (21%)	5 (21%)	2 (8%)	2 (8%)
Female 3	3 (17%)	6 (33%)	6 (33%)	0 (0%)	3 (17%)
Male 2	3 (18%)	5 (24%)	4 (24%)	1 (6%)	4 (24%)
Female 4	4 (27%)	5 (33%)	2 (13%)	1 (7%)	3 (20%)
Male 3*	2 (10%)	8 (38%)	7 (33%)	0 (0%)	4 (19%)
Totals	28 (24%)	35 (30%)	30 (26%)	5 (4%)	19 (16%)

*participant standing rather than sitting

The category with the lowest score, attentional touch, diverged from a higher expected value, most likely because the definition of the category did not correspond to the size and context of the bear. Participants usually held the bear facing in the direction of the object or area of interest in combination with pointing, physically moving the bear when they wanted it to look at something else. As the bear was already in their arms, the need to do something like tapping the bear's shoulder for attention was not necessary, though different behavior would probably be found for a freestanding robot.

There were not enough participants to ensure statistically significant results, but it is interesting to note that the strongest divergences of gender (see table below) were in the categories of Affectionate Touch with an average score of 29% female versus 17% male and Puppeteered Touch, with 23% versus 30%.

Gender	Affectionate	Manipulated	Puppeteered	Attentional	Playful
Female	(29%)	(30%)	(23%)	(4%)	(15%)
Male	(17%)	(28%)	(30%)	(6%)	(18%)

Manipulations of the bear included hugging, rocking, tossing in the air, waving its hands at itself or other people, holding with one arm

sideways/front/out, dancing the bear while holding its paws and laying the bear across their laps.

Over the approximately five minute interactions, seven (four female and three male) participants initiated between 20-50 distinguishable touch gestures toward the bear.

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. Seven out of seven made eye contact with and talked to the bear.

The range of interactions was of course limited by the role-playing situations and sitting/standing position of the participant. In long-term relations with the bear, we would probably also begin to see ways in which users tactilely expressed anger/punishment and acknowledgement/encouragement.

Relevant region sizes ranged from approximately 2x2" in a head pat to a three simultaneous regions in a hug corresponding to the two hands or arms

of the participant and the forward contact between the human and bear. Touch duration and migration were on the order of 2-5 seconds and 5° respectively.

Significant regions began to converge from one participant to the next surprisingly quickly, as can be seen by the following example diagrams.

As can be seen above, 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.

A shortcoming of these experiments was that all subjects were adults.

Children, the main target audience of the Huggable, may or may not have similar touching patterns.

However, it provides a good first estimate, especially in regard to necessary sensor density.

Behavioral Study

Conclusions

The minimum sensor size



Figure 15: Sensor Density and Size

needed to capture the locational content of touch gestures is 2x2", although higher sensor resolution would allow differences in types of touch (e.g. poke, scratch, pet) to be more accurately distinguished.

The minimum query rate for detecting unique touch gestures on the basis of their locational distribution would be 1 second, allowing transition detection and error checking, though behavioral touch will usually require up to 3 seconds for confirmation.

Sensor layout must adequately cover the identified significant regions. A test of one sensor layout scheme on the 3-D bear using sensor-sized paper rectangles taped to the surface of the model bear can be seen on the left.

III-C. Touch Gesture Recognition Design

In this section, I present the design of the current and future bear as regards gesture recognition. The final study included a bear equipped with an array of 58 capacitive sensors that perform pattern recognition to detect symbolic gestures and differentiate touch subtypes. Gesture distributions were characterized using Bayesian nets (that should ultimately use a smoothed Gaussian sensor region model though not included in the current study in the interested of faster classification). Types of local touch are then recognized using the Nearest Neighbors technique.

I also overview the next important steps in furthering and improving the gesture recognition design. Once we have a bear that can recognize a fixed set of symbolic gestures, we should add a variables estimating affective content (angry, bored, content) and appropriateness (gentle, rough) of the touch to better predict what the user behavior is at any one moment. By tracking series of gestures, the bear could estimate what is a user's current interaction mode (i.e. mood and action-plan). These characteristic gesture sequences are used to create an HMM model of user interaction mode given a specified user-bear task of limited duration.

In designing a test exercise that challenged a subject to moderate their physical and expressive actions toward the bear, it will be necessary to run through and verify the validity of each of the touch behavior recognition substeps, though additional high-level testing, a more autonomous bear and meetings with real-life applications and their professionals. One example would be working with behavioral therapists to create a useful system of tasks and interaction to help an autistic child learn and reinforce particular social behaviors.

Touch System Overview

This project makes use a touch gesture recognition model to classify human behaviors and interaction modes, where behaviors are of shorter duration and interaction modes characterize the recent history of behaviors. In a typical pattern recognition system, as presented in the Duda et al textbook *Pattern Classification*, one follows the following sequence: Sensing, Segmentation and Grouping, Feature Extraction, Classification, Post Processing.

In this case, the sensing is an array of 58 capacitive sensors spread over the surface of a foam teddybear, there is segmentation of the data at each time step (running at 3.4 Hz), feature extraction is dependent on the gesture/parameter class, and their extracted features map into our behavioral classification. Post processing allows us to evaluate and improve the system derived thereof.

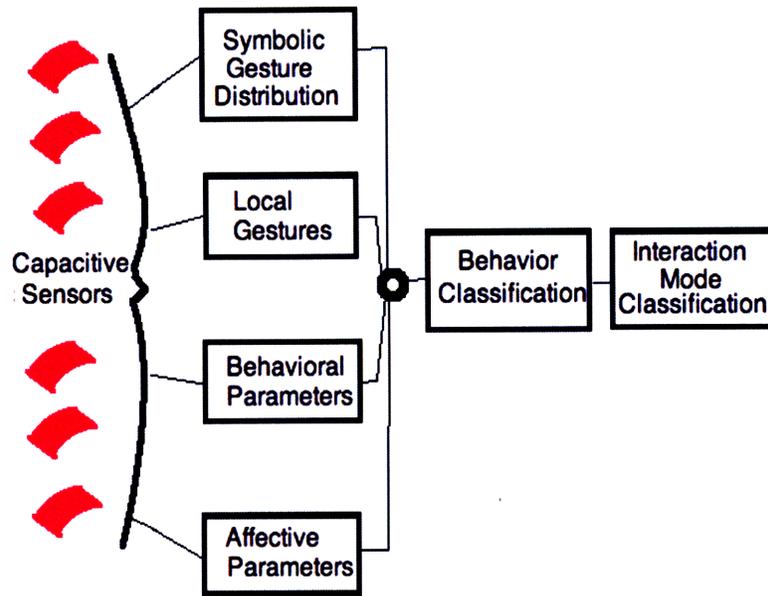


Figure 16: Touch-Gesture Recognition Flow Chart

After coming up with the overall system, the training and evaluation of the models involves: Data Collection, Feature Choice, Model Choice, Training, Evaluation, and consideration of Computational Complexity. In addition to the training on individual gesture and parametric subtypes, this project evaluated the system function through a rescue scenario test evaluation, in which other students in the class were invited to run through a simulated behavioral therapy role-play scenario.

Optimizing Computation Time and Attention

In a system with so many sources of sensor input, it is important to manage data in a manner that is efficient, discarding extraneous details and focusing

on the features with the highest potential relevance. Running a gesture through a regional mapping filters to see if it corresponds to a touch behavior that the bear can recognize is a useful strategy to minimize that computational overload.

Our focus is social communication, thus the bear is set to pay attention to symbolic social gesture distributions exclusively. Once the distribution is classified as having reasonable probability, that hypothesis can be strengthened by the addition of supporting local gesture subtypes.

Compute the easiest computations first, then refine as needed for the task at

hand, evaluating the

probability distribution

for convergence before

deciding to take any

further refinement steps.

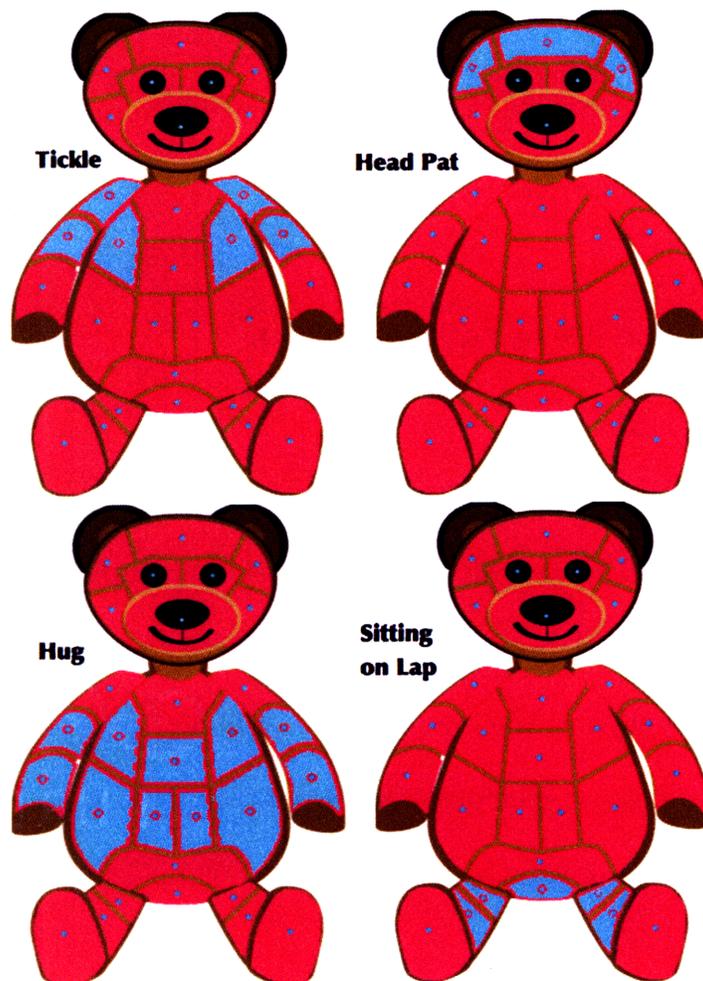
Symbolic Gesture

Distributions

Typical symbolic gestures

included: hugs, head-

pats, tickle, shake awake,



attentional tapping, petting, hand shake, holding hands. Activation diagrams representing a subset of these behavioral gestures follow below:

The bear learns the associated distributions by using a Bayes classifier [14], which generalizes a probability distribution for a current feature based on the history of touch gestures.

$$p(\mathbf{z}|\mathbf{y}, \theta) = \frac{p(\mathbf{y}, \mathbf{z}|\theta)}{p(\mathbf{y}|\theta)} = \frac{p(\mathbf{y}|\mathbf{z}, \theta)p(\mathbf{z}|\theta)}{\int p(\mathbf{y}|\hat{\mathbf{z}}, \theta)p(\hat{\mathbf{z}}|\theta)d\hat{\mathbf{z}}}$$

Then it can use that posterior to make probabilistic inferences about the future. One can use Expectation-Maximization algorithm to learn/improve parameters.

Sociable vs. Procedural Touch

Humans often use touch as a mode of social and affective communication. A social robot needs to distinguish functional from communicative touch. In general, human to robot touch spans the categories: Manipulation, Puppeteering, Attentional, Symbolic. In particular, there are *symbolic gestures*, which have widely used symbolic significance such as hugs or attentional taps on the shoulder. These gestures are characterized by typical

regional distributions, within which particular touch subtypes are present, and have standard duration ranges.

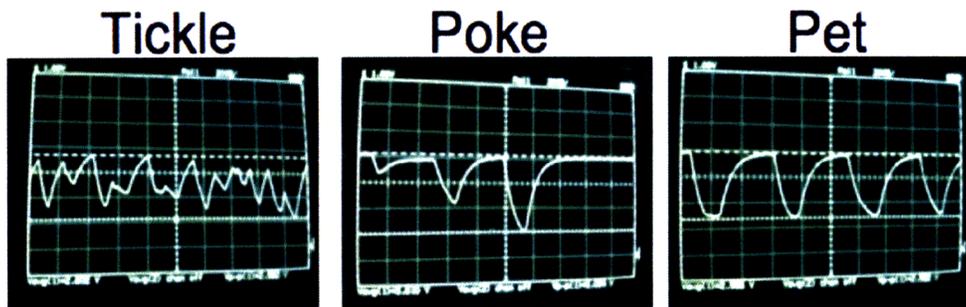
By doing a short behavioral study in which fellow labmates were invited to run through a series of activities with a traditional teddy bear, such as saying hello, waking the bear up, comforting it when it was sad, showing it around the room, I identified a starting set of important behavioral gestures.

Within these gestures were the categories: functional (picking up, supporting or repositioning bear), puppeteering (manipulating the bear such that it responds in socially appropriate ways), and symbolic, with which you are already familiar.

Local Gestures: Classifying Subtypes

The recognition of local gestures, or *touch subtypes*, combines segmenting data at each timestep with the knowledge of previous time sequences such that they update current hypothesis based on the last several seconds. The features used to distinguish these behaviors include amplitude, frequency spectrum, base frequency and duration of the signal at each sensor.

The challenge of classifying time dependent gestures is making them real time, thus I was willing to sacrifice 'perfect characterization' for reasonable but realtime accuracy. Typical local gesture signals are shown below.



Classified subtypes included: Tickle, Poke, Pet, Hold, No Touch

As shown in the following table, different local touch subtypes have distinguishable characteristics that the robot can use to differentiate between them in real time. The data below was taken from direct experimentation with a signal sensor.

	TICKLE	POKE	PET	HOLD	NO TOUCH
Peak Amplitude	60%	>30%	100%	100%	0%
Base Freq	5-10Hz	1Hz / 0Hz	0.5-2Hz	0Hz	0Hz
Freq Spectrum	High noise	Blip	Low Noise	No Noise	No Noise
Duration	3-20 sec	1 sec	>4sec	>4sec	n/a

An additional challenge in the true system, however, is that individual sensor signals are queried at a lower resolution. Because the capacitive sensing mux requires time to acquire signal after switching channels, all sensors are queried in turn with a 5 millisecond delay in between each one.

Thus the true data rate is $(58 \text{ sensors}) \times (.005 \text{ seconds}) = 0.29 \text{ sec}$ or 3.4 Hz, which will eliminate some of the higher frequency information, by the Nyquist rate, to a maximum of about 7 Hz. This has the risk, in particular cases of high frequency tickle of being misclassified as constant touch.

In any case, the specific pattern recognition technique used in this section was nearest neighbors, in which the signal was processed into feature values, which are then classified by into its closest fit touch subtype. Particle filtering would also be well-suited to this domain.

In Psuedocode:

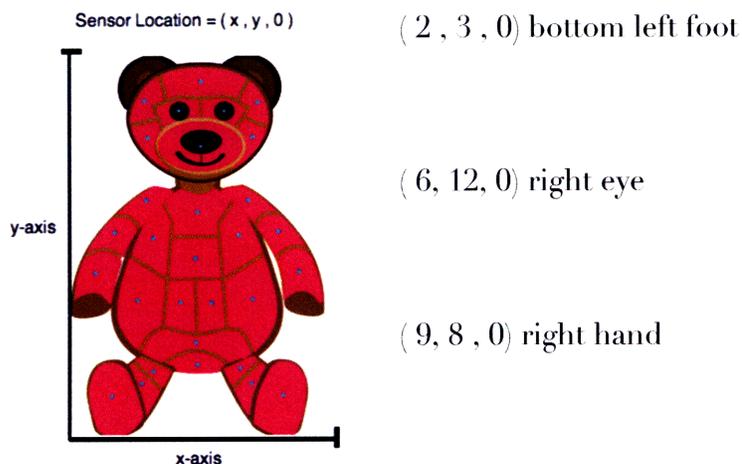
```
At each timestep
If (amplitude < 30%) {
    set_current_class = NoTouch
}
Else{
    Increment duration;
    If (duration < 3 sec)
        set_current_class = Poke;
    Else{
        Update baseFreq, noiseLevel;
        ref_var = Alpha x baseFreq + Beta x noiseLevel
        If (ref_var > tickle_cutoff)
            set_current_class = Tickle;
        else if (ref_var > pet_cutoff)
            set_current_class = Pet;
        else
            set_current_class = Hold;
    }
}
```

Blobbing Gestures: Geometrical Extensions

To learn new symbolic touch gestures or follow a touch that spans multiple sensors in time and/or space independently, the robot needs a way of abstracting away the point sensor information so that it can group contiguous touch over a region.

In order to do so, we must first introduce a geometric representation of the individual sensors for the bear. To simplify calculations in a world where I already had a visualization for the bear, I chose to do the calculations in a simplified 3 dimensions. Thus the z-coordinate simply represents front (0) or back (1) and x,y are the coordinates of the sensor location within the visualization.

Examples:



The back view of the bear, although not shown, shares the same x-y coordinate system as the front thus the bear can cognitively relate gestures that encircle the midsection of the bear or wraparound from the front to the back of the head.



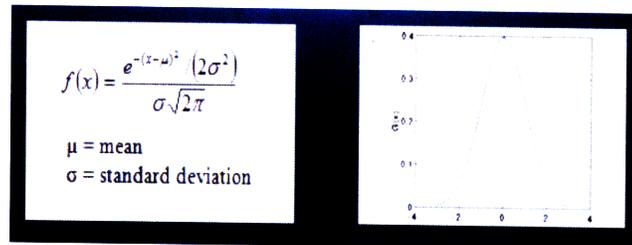
Figure 18: Gaussian Sensors

The inclusion of geometry allows blobbing of touch over several sensors by giving each sensor signal a Gaussian locational mapping. This improves the robot's perception because, first, overlapping regions of the same kind of touch can be cognitively grouped together and, second, traveling gestures can be tracked from one sensor to the next.

Code sensor to center location (μ), then use Gaussian to smooth sensor transitions, acting as a low pass filter. The overlap and frequency range is set by assigning the two board-types (small sensor and large sensor) respective standard deviations.

A Gaussian representation is desirable and convenient as it maximizes information entropy, smoothes the signal, and has the convenient property

that the sum of Gaussians is Gaussian, thus a signal of one or several sensors is represented in the same way.



For moderate overlapping, where the half-width is 2" for the larger sensor and 1" for the small the bear uses $\sigma_{\text{small}} = 0.85$, $\sigma_{\text{big}} = 1.7$, derived from the expression $\sigma = x / (-2 \ln(0.5))^{0.5}$

Thus the regional distribution representation of symbolic gestures can be improved to a continuous blob of varying density, corresponding to point relevance. It is a higher dimension model and closer to truth than the simple point location sensor values.

In cases where one sensor malfunctions, a further property of sampled Gaussians is one can interpolate what the signal should be, as follows: the continuous distribution along a line from x to y is given by convolving the Gaussians at x and y. The result of that convolution is:

$$p(y|x) = G(y; x, \sigma) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(y-x_1)^2}{2\sigma_1^2} - \frac{(y-x_2)^2}{2\sigma_2^2}}$$

The robot could also make use of such a calculation to diagnose the likelihood of sensor failure or evaluate when a new high-interest signal differentiates significantly from the current baseline behavior.

Behavioral/Affective Parameters

Some of the most important, simple and subtle features of this system involve updating the qualitative behavioral and affective parameters for an action. This can be the most critical component to the bear understanding the social/affective meaning of the touch. While a hug is most often a friendly gesture, it never happens the same way twice; it can be affectionate or cursory, tentative or burly.

An illustration of how these parameters can change the meaning a gesture

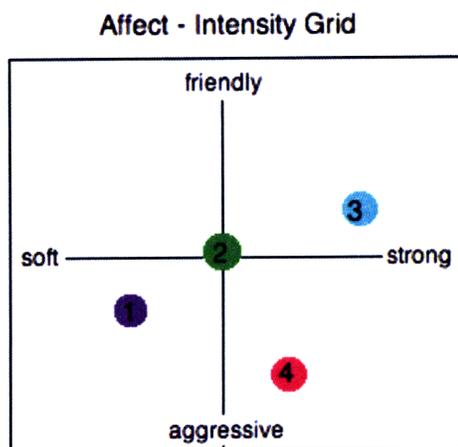


Figure 19: Affect - Intensity Grid

can be conveyed through the Affect-

Intensity graph below. Intensity corresponds to the x-axis of the grid and can run from soft to strong.

Affect runs on the y-axis and spans friendly to aggressive. Let us denote the (x, y) intensity-affect coordinates of a gesture as that gesture's *valence*.

In the case of a punch with a valence of bubble one, the aggressive gesture coupled with the soft intensity could cause the bear to interpret the gesture as more playful than angry, and help the bear characterize the current behavior as rough-housing, or perhaps attentional.

As for the remaining examples, bubble two represents the neutral signal without any additional information from the behavioral/affective parameter. Of high intensity and moderate affection, bubble three could be a hug from a child greeting the bear after a stressful dentist's appointment, looking for comfort but not yet playful. Finally, bubble four is a clearly inappropriate and unfriendly gesture in a world where the bear is a kind agent, and this would be an example of a gesture the bear would discourage the child from doing again.

Encoding Gesture Valence

In this project, the bear encodes these parameters based on the following five illustrative characteristics: low-level features include gesture duration, average gesture amplitude, and affective signature with relevant high-level features being traditional gesture associations and response to feedback.

Intensity is more easily calculated than affect as one can derive it directly from sensor amplitude and duration of gesture. Incorporation of a galvanic sensor that measures the arousal level (intensity of emotion) could more directly measure the cognitive investment the user feels in conjunction with a particular action. Frustration, anger, and gleefulness can all result in similar measurements on the galvanic sensor intensity scale.

Until then, the robot can assess the same measure using signal amplitude and gesture duration. To do that, parameterize each symbolic gesture with a corresponding amplitude coefficient (q) and decay rate (p) as depicted below, where I is intensity, A is amplitude, D is duration.

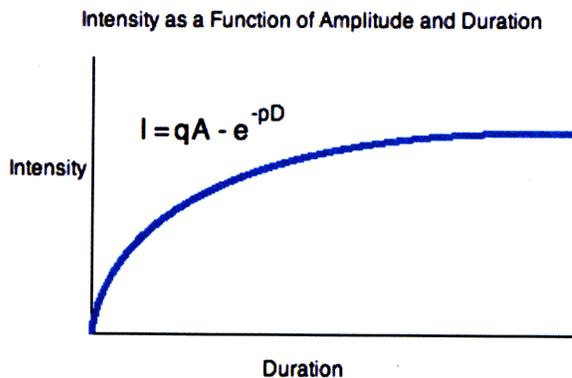


Figure 20: Gesture Intensity over Time

The scaled amplitude sets the maximum value of the intensity function and longer durations increase the bear's assessment of gesture intensity. The reason I chose this representation over linear mapping is that I wanted the signal to converge even for

a long duration signal, and for sufficiently small decay rates, the equation still reduces to a line.

To encode whether the action is friendly or aggressive, the use of ‘affective signature’ would be a useful extension to this project. Unfortunately it requires development of independent characterization infrastructure and training that was outside of the scope of this project. The concept comes from Ros Picard’s book *Affective Computing* [15] in which experiments have shown that even something as simple as a two axis pressure sensor can reliably communicate and characterize the emotional state of the user that is touching it by the signal pattern it leaves. There are physiological reasons why that would happen as emotions change our response times and attention. However, its applications for this context remain to be explored and developed.

Instead, here I rely on traditional symbolic gesture associations as well as whether the users actions are in accordance or conflict with the bear’s feedback. For example if the bear asks someone to please stop tickling him and its requests are consistently ignored what started out as a friendly gesture can later be characterized as an unsympathetic or aggressive gesture. Upon initialization, Hug is very positive, Head pat, Petting and Tickle are positive, Pokes and attentional gestures are affectively neutral, and Slaps and Hits are negative.

This could be extended to slight negative associations with sensitive regions of the bear, such as around its eye cameras or mouth in case we wanted to protect the sensory apparatus of the bear.

Bear Feedback

For the bear's own safety and to serve as a useful tool in behavioral therapy, the bear must be able to give feedback to its user. As a passive sensate bear, this bear's expressive modes are limited to verbal audio response and computer mediated expression of affective reaction.

In these first experiments, the bear itself is audio-puppeteered, that is, a human speaks and listens through the 'mouth' (speaker) and 'ears' (microphones) on the bear, using a visualization/characterization of current touch and a video-relayed image of the interaction to understand subject behavior and decide on the appropriate bear response. In this section, we will refer to the human speaking through the bear as the *remote-operator* and the human interacting directly with the bear as the *user*.

The bear is also capable of giving feedback by communicating its affective reaction using emoticon representations, both locally and for (or by) the remote operator.



Figure 21: Bear Emoticons

In this way, both ends can better understand the bear's current state and its reaction to the user actions. In a semi-autonomous puppeteering mode, the remote operator can use the bear's affective assessment of the interaction before deciding on what his or her next manipulation/vocalization should be.

That operator could also send an emoticon to the user directly, thus puppeteering the affective expression. For example, to discourage a particular mode of interaction the bear would first express surprise or ambivalence, but if the interaction mode continued, the bear would shift to anger or sadness depending on the user's gesture sequence and response.

Ultimately, for use in the real world, a behavioral therapy bear would decide on its verbal and affective response autonomously.

Summary

In my system, there are several independent feature to state mappings from the sensors:

- 1) bodywide \Rightarrow symbolic gesture distribution
- 2) local sensor \Rightarrow gesture subtype
- 3) sensor input \Rightarrow affective signature / valence

After mapping these combined gestures and parameters to user behavior the bear can begin to model human behavior using Hidden Markov Models of relevant user states. In training, the bear can use previous behavior sequences to improve the current behavior estimate such that the bear can provides appropriate feedback when (and only when) needed.

III-D. Software Implementation for Real-Time Recognition

In addition to the physical and electrical design of the Sensate Bear, I also worked with Rob Toscano to create a software interface in Microsoft Robotic Studio. An overview of its function: when started, it loads and initializes a model of the bear sensor layout and activations as well as significant gestures. Next it reads and parses the sensor data streaming in from the serial over USB connection. Then it translates that data into state mappings for the sensor and gesture objects. Finally, it displays this information in the form of a locally hosted website that visualized the bear's currently active sensors and gestures. Rob created the bulk of the code, adapting the website display from a related Huggable program, while I did the

configuration tables, visualization drawings, communications protocol with the microcontroller, gesture classifications and processing.

Sensor Mappings

A configuration file helps the program translate the incoming signal activation levels to sensor locations on the bear visualization. The mappings, shown in diagram below, are set in an config.xml document.

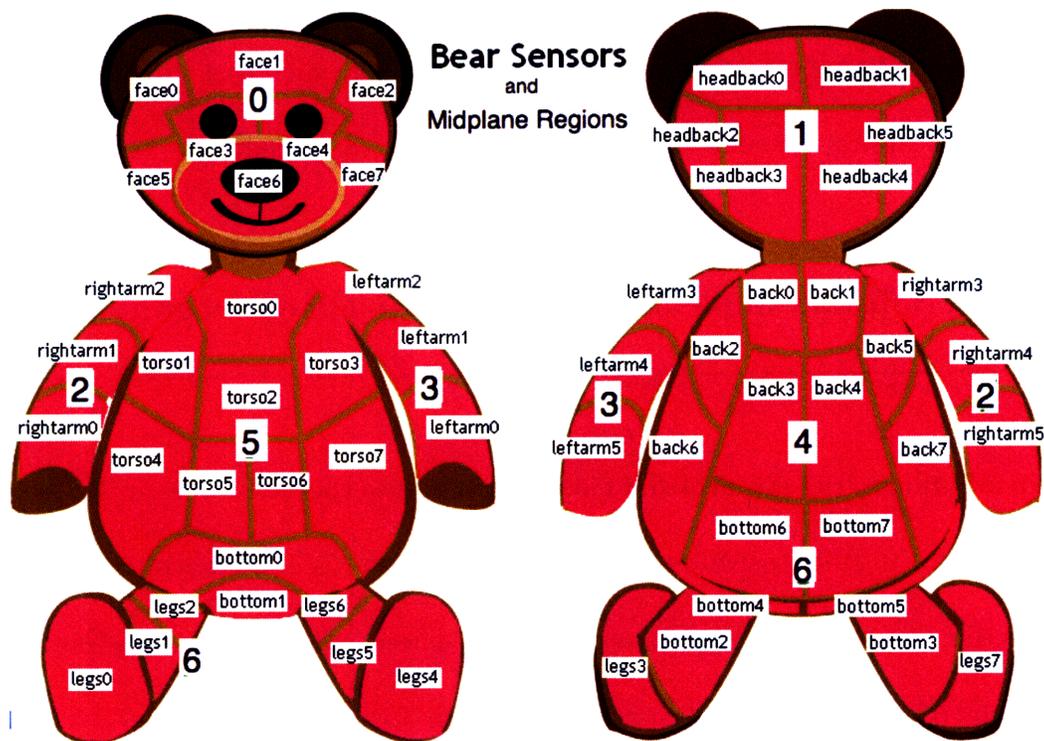


Figure 22: Sensor Mappings for Visualization

Communications Protocol

The microcontroller is set to stream all skin sensor data, in order, as set by the three-byte sensor ID. The sensor ID is the concatenation of body

section value (zero for all sensors), midplane value (can take values 0-7), and sensor channel (0-7), with a byte dedicated to each value.

This format was selected because it is easily expandable to future sensate skin iterations for the Huggable and does not have a preset number of sensors, midplanes and body regions.

An easier-to-read (using sensor number rather than true sensor ID value) sample stream might be:

DATA 0 820 1 820 2 790 3 790.... 56 820 57 820 STOP

Where sensors 0, 1, 56 and 57 had values of 820 or four volts and were inactive and sensors 2 and 3 had value of 790 and were being touched.

To communicate sensor value, each sensor has a corresponding ten-bit number where the highest value (1024) corresponds to 5V and the lowest (0) corresponds 0V. There are also start and stop keys. 'DATA' begins the stream, and 'STOP' finishes it.

Visualization Display

When the incoming sensor signal value is above a certain threshold, the corresponding sensor (displayed in pink) appears on the web interface. There is a constant bear background and then drawing of each sensor with a transparent background. Thus when triggered, the display overlays all active sensor. This sidesteps the need to store sensor location transformations in another part of the program. To further speed refresh and display time, new information is only sent to the website model to turn on and off individual images, rather than reproducing information about elements that are already there.

Gesture Recognition

At each cycle (every time the program reads 'STOP'), the current activation pattern updates its gesture hypothesis. If active, screen text will label any active gesture, as in, for example, the 'Headpat' below.

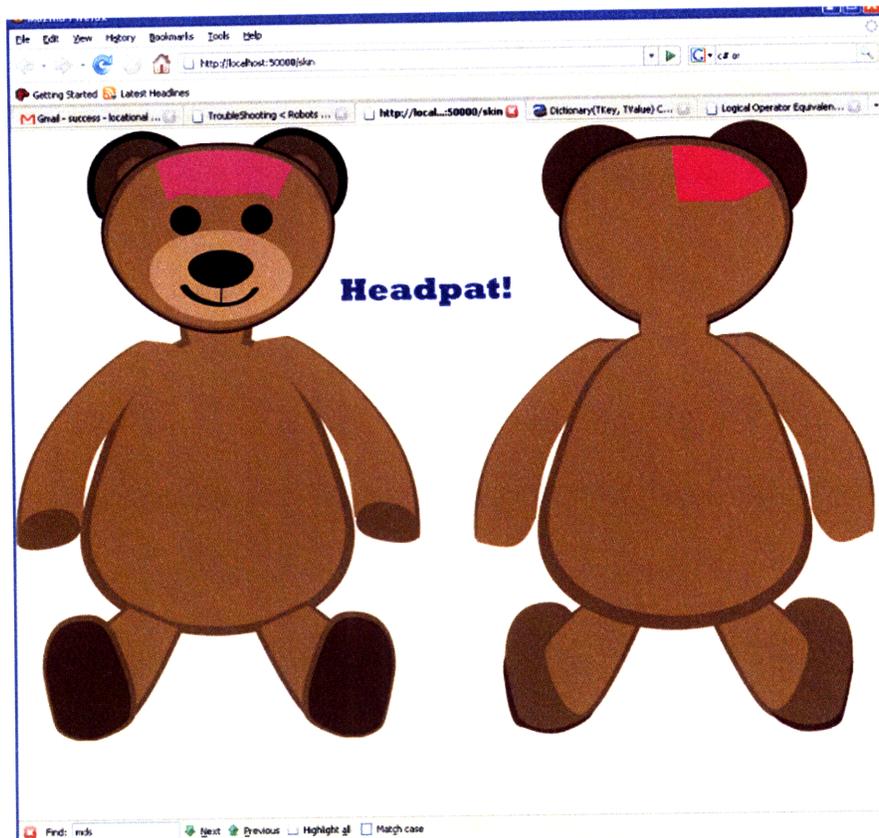


Figure 23: Bear Classifying a Headpat

Labeled interactions for the end study were headpat, hug, tickle, and footrub. The algorithms for each consisted of the adult-observed locational distributions (human pattern recognition) where the minimum number of sensor activations and any additional conditions were met.

Headpat: at least one of the three sensors active

Hug: both sides (of six front, four back) and at least three sensors active

Tickle: not Hug, at least two active of four sensors active

Footrub: at least one of two sensors active

Multiple gestures (with the exception of Hug superceding Tickle) can be active at any one time. These classifications were consistent with the initial behavioral study and represented the most used expressive gestures therein.

Local Subgestures

To better process touch types such as poke, tickle, hold and stroke, I added a variable to Rob's sensor class that can store the last several cycles of values for its activation level. Using that history, we can calculate the relevant feature values, peak amplitude, base frequency, frequency spectrum and duration, as outlined in the gesture recognition design section.

The peak amplitude is the maximum value in the set, base frequency can be calculated from lowest frequency value in the Fourier Transform, frequency spectrum is the band over which the values are within two standard deviations from the maximum, and duration is incremented for each cycle over which a subgesture classification has been the same, resetting for each new gesture.

An interesting test would quantify the efficacy of full spectrum readings, as the simple on-off activation pattern is sufficient to distinguish gesture subtypes. This would provide an additional simplification, sidestepping complex calibration routines in the interest of fast and efficient processing.

III-D. Utility and Acceptance: A Pilot Study with Children

To evaluate and gather further information for the my sensate robot architecture, I invited eleven children to participate in a study with the foam sensate bear. All were under the age of twelve. None hesitated to engage with the bear and their touch, though sometimes shy, was similar to how they would treat a real creature, involving all surfaces of the bear.

The purpose of the study was to evaluate the utility and effectiveness of the Sensate bear, and answer the following questions:

- (1) would children be willing to play with the bear?
- (2) how would they interact the bear, what was their level of engagement?
- (3) did touch play a key role in their expressiveness?
- (4) did the bear have the ability to sense the child?
- (5) were its programmed gestures effective?
- (6) how did seeing the bear recognize gestures change the child's relationship with the bear?
- (7) what additional gestures is it important for the bear to understand?

The study was approved by the MIT Committee on the Use of Humans as Experimental Subjects. I was the study conductor and Angela Chang was the bear's audio-puppeteer.

Study Procedure

Study variable included: presence of audio-puppeteering, visibility of sensor activations, duration of interaction. We also tracked the number of children in the room, gender and age.

Evaluation variables included: eyecontact (bear, study conductor, visualization, parents), initiation of new touch gestures, position of bear (lap, table, in arms), verbal interview of reactions

The bear itself had 58 active sensors, its signals were un-amplified, a time-stamped recording of all sensor data exist for all times when the visualization was active. Study fliers can be found in Appendix B.

Upon the child's arrival the study conductor would follow these steps:

1. Give Participants a 5-minute Tour of Personal Robots
2. Ask Parent to Fill out COUHES Consent form(s)
3. Invite Family to Study Room (my office)
4. Introduce bear, greetings
5. Pass bear to participant [6 through 10 in any order]
6. Left from Right
7. Learning Gestures: Hug, Pat, Tickle, Footrub
8. Open play / discussion

9. Ask Participant to tell bear a story
10. Ask Participant what they would like do with bear
11. Bear and Participant say goodbye
12. Give Participant Huggable Sticker

**video interview and video release form followed for first three participants*

There were variations in the order of the above (particularly in interaction evaluation steps 6-10). Step six was an exercise in which the child was asked to help bear learn its left from its right. Step seven tested pre-programmed gestures, though some characterizations were discovered by the child before this step. In step eight, open play, the children and bear engaged each other or parents and children asked questions about the project in non-scripted ways. Step nine would have been more effective if there was a specific story or storybook the child could have read to the bear, because most did not immediately have a story come to mind. Step ten was an open question to the child.

Chart of Study Participants

:#	Age	Gender	Company
1	10	Male	+1
2	4	Female	+1
3	11	Female	0
4	5	Male	+1
5	10	Female	+1
6	7	Male	0
7	11	Male	0
8	9	Male	+2
9	8	Male	+2
10	10	Female	+2
11	4	Male	0

The sensor data files (c://SkinLogs) can be played back in software for later processing. To do that modify the SkinSensor.state.xml file.

Change the "ImplementationType" from "Serial" to "LogFile", and make sure the "PathToPlaybackFile" element has the right path.

Sensate Bear Study Results

Classifications evaluated: Tickle, Head-pat, Foot-rub, Hug. Head-pat and Foot-rub were consistently identified through all users. Hug was sometimes identified, but using the unamplified signal meant that the sensors did not detect contact with clothing. Looking at the video of what the physical contact locations were, that classification would probably be accurate with calibration. 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.

Summary of additional symbolic gestures that the children expected the bear to understand: Hand-shake, Belly-tickle, Back-scratch, Foot-tickle, Shake-awake, Go-to-sleep, Lay-to-sit, Feeding, Rocking.

Hand-shake: grasp bottom/middle arm sensors and move arm up and down.

Belly-tickle: tickle sub-gesture over protruding midsection of belly

Foot-tickle: tickle sub-gesture on bottom of foot

Shake-awake: grasp shoulders or midsection and shake bear

Go-to-Sleep: put bear in laying down position to go to sleep (children suggested also indicating naptime to the bear by covering it with a blanket)

Lay-to-sit: another method several children used to wake the bear was moving it from a lying down to sitting position

Feeding: putting play food in the bear's mouth

Rocking: cradle or hold bear with planar horizontal rotation

Bear Positions (from most to least common): Sitting on lap, Sitting on table in front of child, Held in arms, Lying on lap, Lying on table, Held in air over table.

Bear Manipulations: Pick-up, Sit back on table, Make bear dance

Chart of Study Variables

Participants	Visualization	Audio/Video Recording	Audio Puppeteering
I, II	Yes	Just Audio	Yes
III	Yes	Partial	Yes
IV	Yes	Yes	Yes
V	Yes	Yes	Yes
VI	First Half	Yes	Second Half
VII	No	Yes	Yes
VIII,IX,X	No	Yes	Yes
XI	Yes	Yes	Second Half

Discussion

Returning to the motivating questions:

(1) would children be willing to play with the bear?

Yes. Eleven out of Eleven did not hesitate.

(2) how would they interact the bear, what was their level of engagement?

Gestures mirrored those used toward social creatures and included those outlined in the discussion, as well as grazing movements touching or moving the arms and legs while talking to other people in the room, much like doodling while on the phone. To measure engagement I looked at eye-contact, conversation partners and numbers of new touch gestures.

(3) what role did touch play in their expressiveness?

Consistent with the results of the adult behavioral study, much of the touch expressed affection or occurred during manipulation of the bear. The interactivity of the bear provoked new kinds exploratory touch as they tested the capability limits and response of the bear.

(4) did the bear have the ability to sense the child?

During the right vs. left section of the study we tested the base ability of our system to sense a child. All children were successful at triggering the appropriate sensors, seven out of eight were successful on the first try. Several children were confused about right vs. left themselves, perhaps that made them better able to relate to the bear. Note: this study was performed without amplification, so the bear could only sense skin contact.

(5) were its programmed gestures effective?

See discussion. Footrub and Headpat were always sensed, Hugs 40% of the time, 60% with explanation, and Tickle 20% first time, 80% with explanation. Needs more training and signal calibration.

(6) how did seeing the bear recognize gestures change the child's relationship with the bear?

Children would laugh, point at the screen, draw their parents attention to the gesture, perform it again. Audio-puppeteering provided a constant social response to any gesture, so visible classifications in this mode provoked less response.

(7) what additional gestures is it important for the bear to understand?

See discussion. Hand-shake, Belly-tickle, Back-scratch, Foot-tickle, Shake-awake, Go-to-sleep, Lay-to-sit, Rocking.

The children's interaction also altered with changing study variables and there were several unexpected emergent traits, such as parent and sibling dynamics and shyness.

Culture and Personality

The sample size was limited, but variability in social mores within the eleven test subjects indicate an influence of culture and personality on a child's readiness to be affectionate, expressive or aggressive with the bear.

The Interaction Warm-Up Period

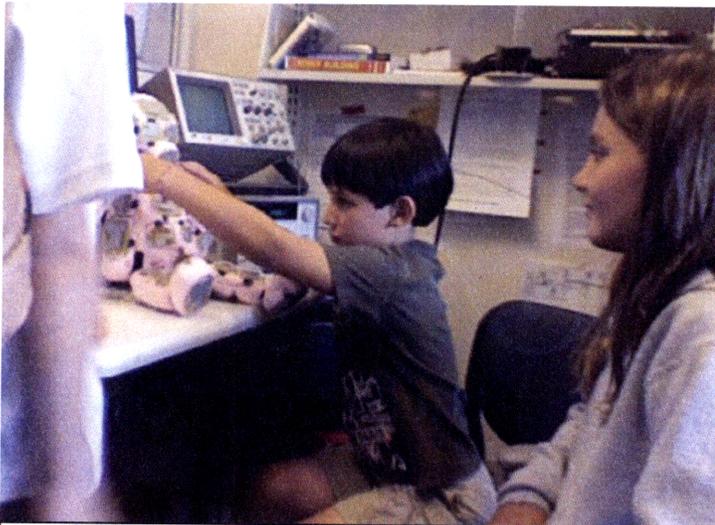
A emergent trait not present in the adult behavioral study was a child's (particularly when alone) initial shyness with the bear. There was a pronounced warm-up period on the part of the child as soon as the audio-puppeteering was active, in which its very first estimates can be accessed remotely. The study was short that the results will likely be different as they become more comfortable with the bear. They might take longer, but once hooked, they will probably open up to the bear significantly faster than most adults.

The Effect of Age

The older the child, the faster they were to understand the tasks and be open with the bear. The older children also engaged in more social touch gestures, while the youngest were most fascinated with the on-off reactions of the visualization.

Group Dynamics: Interaction en Mass

Parents were always in the room, helping the children understand the robot



or suggest new interactions and there was also a large effect on interaction style was when there was more than one child in the room.

Figure 24: Children with Bear

En mass, sibling groups helped each other better understand how the bear worked, encouraging or 'egging each other on' to try things they might otherwise take longer to do, whether a simple hug passed around, or more mischevous plots to trick the audio-puppeteer.

This context provided a naturally social environment, in which the bear

interactions were a function of the child, study conductor, parents, bear and audio-puppeteer.

Audio-Puppeteering

All children played along with the the audio puppeteering, even though they all seemed to also realize (I told them if they asked) that Angela was outside the room, speaking for the bear.

The audio source was not at the bear but at the computer videotaping the interactions. That usually caused some initial confusion and eyecontact sometimes went to the computer rather than the bear, but after the first minute, the child seemed to accept that it was the bear itself speaking.

Interactivity

With audio-puppeteering, the children conversed with the bear fluidly and responded to the bears requests or suggestions of touch with that gesture.

More procedural conversation (e.g greetings, exchanging names, informational questions) did not provoke many touch interactions.

However, sound effects like the bear falling asleep or laughing did. These tendencies seem to 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). The above should be verified in a followup study, specially designed with experiments evaluating those hypothesis.

In the case of the visualization, subjects were consistently engaged in testing the functionality and reactions of the bear. When inactive they were less successful at activating sensors or gestures but when it was there, they learned those skills quickly and seemed to enjoy the visual response, particularly among the younger audience. 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.

Operating with both the visualization and audio-puppeteering at the same time, either made the child 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. A better coordination of verbal and visual response would better mesh the two modes.

IV. Discussion and Conclusions

As shown in the final study, we have demonstrated a system capable of understanding realtime touch from untrained users, learning that children readily engage in touch interactions with the bear and express hope for increased interactivity in the future.

We have an appropriate sensor density for social gestures, the bear will only need the addition of joint potentiometers and an inertial measurement unit to complete detection. The final architecture block diagram is shown below.

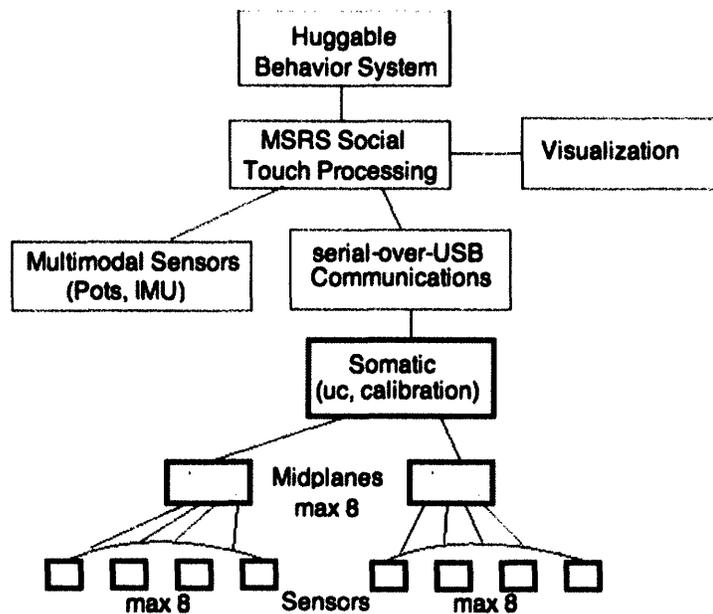


Figure 25: Touch Architecture Block Diagram

All of this work will be funnelled into a Touch API for robots, that will be an extension of the software created for this thesis. The Huggable behavior system will be able to query the Touch Module for currently active sensors

and values, gestures and subgestures.

The capacitive sensors proved highly capable of social gesture recognition. A low density network of on-off sensors provides rich information about current gestures. The biggest problems with large sensor arrays is, first, wiring and second, processing the data in real time. The techniques outlined and implemented here, however, are able to do that. Adding calibration, training off the data sets and incorporating subgesture classifications will complete the first generation robotic touch architecture, which can then be put immediately into use on the real bear.

V. Future Work: Active Touch and Multi-Modal Sensing

As mentioned above, further studies and the inclusion of an inertial measurement unit (IMU) and potentiometers in the joints of the bear would greatly improve the robot's gesture recognition capabilities. Additionally, evaluating a bear with active touch that is more locally expressive will likely provoke new behaviors from its human companions that will need to be newly stored and classified.

IMU and potentiometer technologies have been separately implemented in other iterations of the bear, and what follows is a strategy to combine them together.

The challenge in having multiple sensor modalities is how to combine the data in an efficient way to make accurate predictions about what is happening. I propose again to use an Gaussian model to estimate the likelihood of the current readings, extracting features from each sensing modality related to desired gesture knowledge, weighed by an index of each sensor's reliability. To optimize sensor computation time, we can query the most relevant and low-cost sensors first, only continuing on to evaluate other sensors if the probability thresholds are not sufficiently high for classification.

Further, we can direct attention to particular sensor regions depending on robot state and the task at hand. Finally, the robot can use human behavior modeling in the context of its niche activities to better understand what that task might be.

Additional Training

The final study was conducted with a passive sensate bear with touch/no-touch sensor activations. Spending more time to train our system off the study results would extend the number of gestures the bear can recognize. Next we should acquire new data from calibrated spectrum sensors (make use of full 10-bit sensor reading) with the fur on. This would help hone and

improve the accuracy of the gesture recognition system, and is critical for the local subgesture feature extraction.

Evaluate a Robot Capable of Active Touch

The next major step is to equip the test-bear with active touch by installing the shoulder motors, thus the robot will be able to respond to and initiate touch. Passive touch is an important first step, but it is also important to study how active touch affects social engagement. Though there was audio and visible sensor activations in the final study, our hypothesis is that a self-moving robot would ensure fewer one-sided interactions, better expressing feedback and confirmation of response. Thus, a machine that can initiate and receive touch will likely generate a different sweep of data to be incorporated into our analysis.

Processing a Single Sensor for Multimodal Analysis

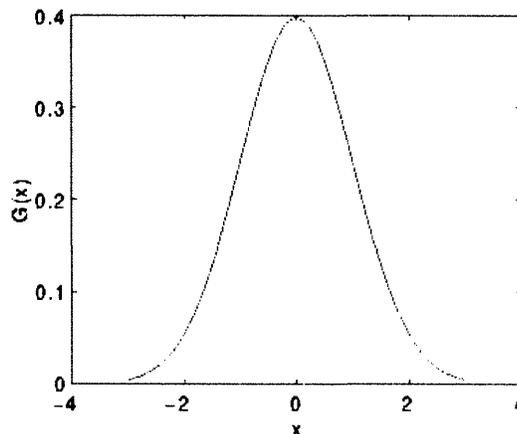
Using a Gaussian to map an individual sensor to feature provides a reliable and flexible model for sensor behavior. To do so, we take the sensor reading as the amplitude at the center of the sensor and use a probability density function with normal distribution to interpolate evidence values at other locations.

The normal distribution is one of the most widely used statistical distributions and maximizes information entropy among all distributions with known mean and variance. Thus we can characterize the sensor reading as the mean and the variance is given by the characteristic error of the device.

$$f(x) = \frac{e^{-(x-\mu)^2 / (2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

μ = mean

σ = standard deviation



For example, if a laser range finder reads 10 meters, however we know that this particular model has tends to have an error of 10 cm, then the mean is 10m and the variance is 0.1 (10%).

Given a particular sensor reading (s), what is the likelihood of a particular feature (y), the $P(y | s)$? Assuming a high reliability sensor, the probability of a the true human position being at y degrees is given by $f(y)$, where the Gaussian mean is the reading and the the deviation is a constant for the device. Thus, $P(y | S) = f(y)/$

The evidence given by a sensor/detection algorithm for a particular feature depends on the current sensor reading, the error of the sensor, and the

feature condition we are testing.

Another index that we must consider is the sensor reliability. In other words, given that a sensor concludes X, what is the probability that X exists, not that X exists at that point, but that X exists at all. How often will there be false readings? We can measure this experimentally using false positive rate, where:

T - True reading, F - False reading

D- Feature Detected, ND- No Feature Detected

$p(T | D)$ = Detection Reliability

$p(F | D)$ = False Positive Rate

$p(T | ND)$ = Missed Detection

$p(F | ND)$ = Detection of Absence Reliability

Depending on the task at hand, different aspects of the sensor reliability may be more important. For example, if a robot is trying to detect a human fall, it is much more important that it not miss any true positives, even if there are occasionally false positives, because a fall could be life-threatening. However, if we are detecting whether a human wants to be assisted finding a product in a store, it is probably better to miss a few customers whose interest level seems borderline and focus on those clearly interested in interacting with the robot, than hound customers that may then find the robot extraordinarily irritating.

Interpreting Multiple and Multimodal Sensors

To expand this strategy to many sensors, one can make use of some handy properties of Gaussian distributions. In probability theory, normal distributions arise as the limiting distributions of several continuous and discrete families of distributions. So our strategy for fusing multimodal sensor readings is straightforward: each sensor will give a weighted prediction about some feature. Sum predictions.

We're already using Gaussian probability distributions to model our sensor readings, and fortunately the sum of Gaussian predictions for a particular feature happens to be Gaussian. In fact, the sums of large numbers of non-Gaussian random variables with finite variance are Gaussian too. This characteristic is one of the central properties of probability theory, the Central Limit Theorem.

Central Limit Theorem: as the sample size n increases, the distribution of the sample average approaches the normal distribution with a mean μ and variance σ irrespective of the shape of the original distribution

Take the array of capacitive sensors on the bear's stomach. More than one sensor is likely to be activated if a person's hand is touching a

particular area, however each sensor will have a slightly different distribution predicting the distance away or strength of touch. However, at the limit, they will sum to a Gaussian prediction of where and how intense that touch is. [14]

For example, try to guess a distance feature from a variety of distance measuring instruments and methods, in a case evaluating the likelihood that there is a person at $z = 5$ meters. Model sensor data as a sum of k Gaussians each weighted by its probability, where the probability weights sum to one.

$$p(z) = \sum_{k=1}^K \alpha_k G(z; \mu_k, \sigma_k); \quad \sum_{k=1}^K \alpha_k = 1$$

Look at the evidence for a feature z (a person is 5 meters away), by looking at the evidence given by each function individually, the value of each individual Gaussian for 'z.' Scale the relative contribution of each sensor by assigning weights (the laser range finder might be three times as reliable as vision triangulation, so their weights might be .75 and .25 respectively). The resulting sum will provide the probability, or evidence, for a particular feature z .

Optimizing Sensor to Feature Computation Time

In order to optimize sensor processing, we can simplify continuous feature calculations, look to the least computationally intensive sensors first and direct the robot attention to the most relevant and salient regions.

In the case where you are interested in a continuous probability distribution between more than one feature values, Gaussians make it easy. If you have already preformed calculations for more than one feature value, say both 5 meters and 6 meters, the continuous distribution between those values will be the convolution of their independent distributions.

$$p(y|x) = G(y, x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}}$$

Thus, our model can represent probability information for both discrete and continuous features and our Gaussian model can save you calculation time by reducing the number of unique computations necessary at an individual point [16].

Another future strategy to avoid state space explosion when the skin is integrated with the various other sensors and robot inputs (vision, IMU, joint potentiometers) is to prioritize the sensor computation order such that the inexpensive computations are done first, and you only proceed to compute the rest if there was no local maxima after the initial computation.

This is not an idea that should be applied to all scenarios. However, when

developing applications that needs a rapid exploration of space and can deal with the fine tuning later, it is a brilliant solution. If someone trying to get the attention of the robot taps him on the shoulder, the robot should first make use of the touch location to rapidly find its partner, and then focus on tracking them with vision. The touch gives a great ballpark estimate of the person location, and for initial movements is probably the only modality needed, but as the orientation gets closer to the person, the robot will need to use the camera to achieve a clearer local maxima for the final video framing positioning.

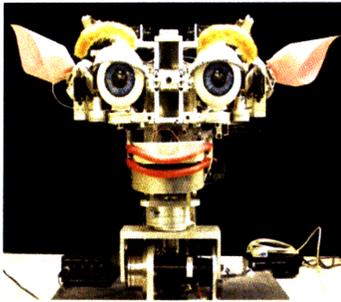


Figure 26: Kismet

Finally, depending on what the robot is interested in we can reduce the overall number of computations by focusing the robots attention on the relevant information. Look at the task set for the robot and state of the people around it. If the store assistant robot already has a customer, it shouldn't waste its time looking for more.

In the case of a social robot, the design of its attention system can reduce computation time, as was created for the robot Kismet [17]. Kismet has a fairly simple behavior system, its sensors consist of a face-tracker, a motion-tracker, a saturated-color tracker and it also has a tendency to get bored and its tasks are to socialize with people or to play with toys.

The sensor data is added together with varying individual weights to decide what is most interesting to the robot at that particular moment. In this case, the weights encode the robot's attention state at that time. When the robot is ready to socialize, the color detector weight might be zero, so the robot could skip the calculation and ignore toys altogether, but the face detector would be center stage.

Simplified Psuedocode Attention Scheme:

```
enum drive = [socialize, play]
while(1){
  while(not bored){
    if (drive = play)
      look for saturated colors
    if(drive = socialize)
      look for faces
  }
  switchDrive()
  reset boredomTimer
}
```

The robot can have two drives, socialize or play, that drive lasts until the robot gets bored. So, while the robot is not bored, if the drive is to play, the robot should just look for saturated colored, but if the drive is to socialize it should look for faces. Otherwise it should switch its drive and reset the boredom timer. A true attention system might be more complicated than this, it captures the concept of how to limit computational attention, given a particular robot task.

Hidden Markov Models: Identifying Human-Behavior State

Sensors are the only inputs that a robot has to perceive outside reality, so the choices of what sensors we use and how we combine them is the only way to get information into the Hybrid Markov Models.

HMMs are good at providing a higher level abstract understanding of what a person seems to be doing, but to detect that a behavior is changing, it needs to map its sensor information into perceptive information, aka features, that fit into the higher level model [18].

The state is 'hidden' because only the human knows his or her own intention, so any classification that a robot makes is by definition, a guess. This model assumes current state can be predicted from past.

Related Methods:

- estimate state with belief state update
- estimate state history with Viterbi
- parameterize an HMM with Baum-Welch or Expectation Maximization (EM)

Modeling Human Behavior

Building on HMMs, we can abstract behaviors into particular sequences of states to make them easier to detect. This can be done in parallel with the general HMM state model. One example of how to do this is to identify states that indicate the possibility a particular behavior is starting, then intermediate states that are necessary substeps within the behavior. In this way, one can look for features that confirm or contradict a behavior, in which case the supposition will be kept or disregarded.

This technique is borrowed from a robotic soccer project [19]. In this research, potential robot behaviors fell into predefined set, i.e., going to goal, getting behind ball, blocking another robot. The purpose of the algorithm, the Behavioral Hidden Markov Model is to allow one of the robots to predict which of those behaviors another robot might be currently doing.

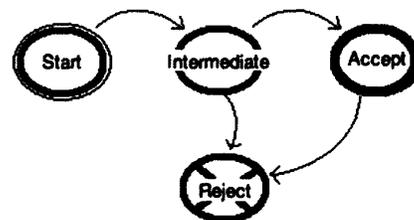


Figure 27: Behavioral HMM

The notation they use includes, start, intermediate, accept and reject states.

The observation values are used to evaluate the current state at each time

step. If the sequence follows a predefined behavior pattern as given by such graphs, the robot will conclude that that is the action.

The basic algorithm is as follows (keep in mind we are segmenting data at

each timestep as well as maintaining multiple hypotheses). At each timestep, if there is a new possible behavior, add it to the hypothesis list, check all current hypothesis and update their states, if there are reject states, remove hypothesis, if a behavior is in the accept state the behavior hypothesis.

Technique Synthesis: Multimodal Sensing and Human Behavior Modeling

One of the key points is to layer the abstractions. For example, when deciding on states, only features matter, and when deciding behaviors just consider states. To prioritize which sensors to read, consider the attentional demands of task at hand, and after processing, all that's left of the original sensor values are the abstracted feature parameters.

The robot's sense of both the state and behavioral context informs the robot's decision on what personal task or behavior mode to move into. That local mode will affect the attention of the robot, and it should choose the sensor modes most relevant to that status for computation, computing only the low cost sensor modalities initially if first estimates are sufficient. After computing relevant features, use HMMs to determine the most likely current state given previously gathered probability distributions, and finally, if relevant, sequence those states into human behavior hypothesis.

VI. Acknowledgements

Cynthia Breazeal for her pioneering work in the field, phenomenal research group and ability to crystalize half-sentence solutions that change my approach to a design, analysis or user studies for months afterwards.

Walter Dan Stiehl for the conception of Sensate Skin, the Huggable project, invaluable debugging/compiling help, and for being the one who initially suggested I be an undergraduate researcher in the lab way back in 2002.

Yi Wang for her long hours of Eagle, battling with the Calibration circuitry and broken Op-Amps with irrepressible persistence.

Robert Toscano for the crash course in Microsoft Robotic Studio, my visualization, and playback file advice from an internet cafes while hiking in the mountains.

Angela Chang for her assistance in the final pilot study as charming audio-puppeteer.

Adam Harvey for the visualization bear illustration, encouragement and adorable bear emoticons.

Sigurður Örn for the late night microcontroller sessions and endless good spirits.

Jason Carver for our joint research on sensor fusion and human behavior modeling in our Cognitive Robotics course.

Fardad Faridi for drawing the stickers that we gave away to the children in my final study.

Jesse Gray and Guy Hoffman, my glorious office mates who put up with my furry foamy piles of electronics.

Polly Guggenheim for keeping everything running, from Digikey orders to Wednesday group lunches.

Mikey Siegel, Jun Ki Lee, Philipp Robbel, Jeff Lieberman, Matt Berlin, Cory Kidd, John Mcbean, Matt Hancher, Zoz Brooks, Andrea Lockerd Thomas, Blake Brasher, Josh Strickon, Ron Roscoe, Mary Udoji, Tish Callanan and Charles Knight.

VII. References

- [1] Stiehl, W. (2005) *Sensitive skins and somatic processing for affective and sociable robots based upon a somatic alphabet approach*. S.M. Thesis in Media Arts and Science, MIT.
- [2] Kidd, C. (2007) *Designing for a Long-Term Human-Robot Interaction and Application to Weight Loss*. Thesis in Media Arts and Science, PhD.
- [3] Breazeal, C. (2002) *Designing Sociable Robots*. MIT Press.
- [4] Stiehl, W. & Breazeal, C. (2005) *Design of a Therapeutic Robotic Companion for Relational, Affective Touch*. In Proceedings of Fourteenth IEEE Workshop on Robot and Human Interactive Communication (Ro-Man-05), Nashville, TN. 408-415.
- [5] Hoffman, G. (2007) *Ensemble: Fluency and Embodiment for Robots Acting with Humans*. Ph.D. Media Arts and Sciences, MIT.
- [6] McLuhan, M. & Lapham, L. (1994) *Understanding Media: The Extensions of Man*. The MIT Press.
- [7] Sherwyn, D., Lynn, M. & Le, J. (1998) *Reach Out and Touch Your Customers*. Cornell Hotel & Restaurant Administration Quarterly, June 1.
- [8] Goodwin, Fred (host). (2002) *Sense of Touch*. The Infinite Mind, Public Radio Show. Lichtenstein Creative Media. October 25.
- [9] Field, T. (2001) *Touch*. MIT Press.
- [10] Parshall, D. (2003) *Research and reflection: animal-assisted therapy in mental health settings. (Issues and Insights)* Counseling and Values. Oct 3.
- [11] 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.
- [12] Inman, M. (2007) *Giggling Robot becomes on of the Kids*. New Scientist, November 5.
- [13] Yohanan, S. et al. (2005) *Hapticat: Exploration of Affective Touch*. In Proceedings of the Seventh International Conference on Multimodal Interfaces, Affective interaction session. 222-229.
- [14] Driver E. and Morrell D. (1995) *Implementation of continuous Bayesian*

networks using sums of weighted Gaussians. In Proc. Uncertainty in AI, pages 134-140.

[15] Picard R. (1997). *Affective Computing*. The MIT Press.

[16] Sherrah J. and Gong S. (2001) *Continuous global evidence-based Bayesian modality fusion for simultaneous tracking of multiple objects.* Proc. IEEE Int. Conf. Computer Vision, pp. 42-49.

[17] Breazeal C. and Scassellati B. (1999) *A context-dependent attention system for a social robot,* Int'l Joint Conferences on AI.

[18] Tran, D.T.; Phung, D.Q.; Bui, H.H.; Venkatesh, S., (2005) *Factored State-Abstract Hidden Markov Models for Activity Recognition Using Pervasive Multi-modal Sensors.* Intelligent Sensors, Sensor Networks and Information Processing Conference, 2005. Proceedings of the 2005 International Conference on , vol., no., pp. 331-336.

[19] Han K. and M. Veloso M. (1999) *Automated robot behavior recognition applied to robotic soccer.* Proceeding of the 9th International Symposium of Robotics Research, pages 199-204.

Appendix A. Behavioral Study Script

Afternoon With Bear (January 16, 2008)

I. Greetings:

Today I'd like to introduce you to Bear!

I'll let you two say hello and then we'll do some exercises together.

I think Bear's made a new friend, can you introduce Bear to me?

II. Role-play:

I'm going to tell you what Bear is 'doing' and you respond appropriately

1-Bear's tired

2-Bear has an itch he can't reach

3-Bear's excited

4-Bear wants to have a dance party

5-Bear wants a hug

6-Bear can talk

7-Bear's ticklish

8-Bear fell asleep

9-Bear is teaching you quantum physics

10-Bear demands to have a cup of tea

11-Bear wants a tour of the lab

III. Open session:

Is there anything else you'd like to try with Bear?

IV. Closing:

All right, I think it's time for Bear's naptime!

Thanks for your help!

Appendix B. Sensate Bear Study Materials

Flier 1:

Sensate Teddybear seeks Playmates!

Research by Heather Knight, Masters Candidate
Personal Robotics Group, MIT Media Lab, E15-468

Let me take your kids off your hands for a 10-minute study, Tuesday, August 19, 2008! It will take place on the 4th floor of the Media Lab, room E15-468 (take a left when you get off the elevator).

The purpose of the study is to explore the gestures children use to play and communicate with a teddybear. We will use the results to design the haptic systems of the Huggable, a robotic teddybear. The ideal ages are 4-12 years old, and participants will take turns to play with the passive sensate bear.

A study guide will be in the room with the child to explain the instructions for each segment of the COUHES approved study, which will include roleplay, requests for specific gestures (e.g. hug, head pat), and a short verbal survey of their reactions. The bear itself will not move, having instead a 'puppeteered' voice from an operator in the next room. In each experiment, we will store logs of the touch sensor and video data in order to better label the interactions.

All participants will receive a Huggable sticker and a brief tour of the Personal Robots group's robots! Please forward this on to others that might be interested.

Flier 2:

SPECIAL ROBOT TOUR FOR KIDS

MIT MEDIA LAB

TODAY - Lunch 11:30-2 and Afterwork 4-7pm - TODAY

Bring your kid(s) to play with a passive Sensate Teddybear for 10 minutes and they'll get stickers and a brief tour of the Personal Robots group! Just call and join us on the 4th floor of the Media Lab, room E15-468 (take a left when you get off the elevator). Recommended ages 4-12 years old. The interaction study is COUHES approved.

More details at

http://web.mit.edu/rehtaeh/Public/SensateBear_Flier.pdf