Real-time social touch gesture recognition for sensate robots

The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Real-time Social Touch Gesture Recognition for Sensate Robots

Heather Knight*, Robert Toscano*, Walter D. Stiehl*, Angela Chang*, Yi Wang, Cynthia Breazeal*,
**Members, IEEE*

Abstract—**This paper describes the hardware and algorithms for a realtime social touch gesture recognition system. Early experiments involve a Sensate Bear test-rig with full body touch sensing, sensor visualization and gesture recognition capabilities. Algorithms are based on real humans interacting with a plush bear. In developing a preliminary gesture library with thirteen Symbolic Gestures and eight Touch Subtypes, we have taken the first steps toward a Robotic Touch API, showing that the Huggable robot behavior system will be able to stream currently active sensors to detect regional social gestures and local sub-gestures in realtime. The system demonstrates the infrastructure to detect three types of touching: social touch, local touch, and sensor-level touch.**

I. Introduction

The physical nature of robots necessarily dictates that detecting different levels of touch is an important area of research. As robots become social actors with the ability to physically engage human bodies, we must develop a social touch taxonomy to describe the new realms of interaction. *Social touch* is defined as touch that contains social value.

Touch plays a key role in human interaction [1]. Waitresses who unobtrusively touch their customers receive larger tips [2] and babies need touch for healthy, balanced development [1]. Prior work demonstrates that a robot which can respond to affective displays (such as petting in pet therapy [3][4], or in playing with children [5][6]) can increase sociability, reduce stress levels and evoke social responses from people.

Touch researchers have demonstrated that the social value of touch varies depending on the body location and various other factors (duration, relationship of people) [7]. Our work looks to populate a social touch taxonomy by observing social gestures demonstrated by humans, and implementing pattern recognition. The closest algorithmic match is [8], which uses both region and manner of touch to automatically group common clusters of touch.

We note that much of the prior work falls into detecting sensor-level touch rather than detecting the symbolic value of touch. Pressure-based robotic skin systems include [9] and [10]. We apply similar algorithmic strategies for local touch to capacitive sensing, which senses both human contact and close proximity touch.

Our hypothesis is that a body-awareness of touch, combined with the gesture profile of the touch, can allow a robot to detect a socially laden gesture (like a hug) as well as local gestures (like a poke). This creates a system that can infer social meaning from the contact between a human and a teddy-bear body.

We define *sensor-level touch* as the robot's knowledge of the activation and location of each individual sensor. *Local touch* allows single-sensor discrimination of a tickle from a poke. Prior work has demonstrated the detection of *local touch* sub-gestures using tactile resolution and gesture profiles to detect affective content. In this work, we attach a social value to touch at different body locations to determine *symbolic touch,* which posits that there is a locational significance to touching an anthropomorphically shaped robot body.

This paper describes our development of a system of real-time touch classification for full body touch sensors. We track gestures across the geometry of a teddy bear using an initial touch gesture library gleaned from behavioral studies with adults. In ongoing work, the sensor system and architecture presented in this paper are being incorporated into the Huggable robotic teddy bear [11], [12].

II. Robotic Platform

A. The Huggable

The immediate application for this research is to equip the Huggable personal robot platform with sensate touch. It will use touch to better perform its roles in healthcare, social communication, education and entertainment applications [13]. In prior work with the Huggable, we classified a diverse set of affective touch interactions for a paw segment with pressure, temperature and capacitive sensors using off-line techniques [11].

B. The Sensate Bear

In order to further develop the tactile taxonomy, we created a separate hardware test system for stand-alone touch processing. This platform is called the Sensate Bear, and has a lower-density of electric field sensors spread throughout a teddy bear body [14]. The creation of a separate rig allowed us to make our somatic processing system full-body and real time while the current 3rd generation Huggable robot was still being developed and built. The algorithms developed with the Sensate Bear test system will be used in the 3rd generation Huggable robot for real time affective and social touch processing. Figure 1

Manuscript received March 1, 2009. This work is funded in part by a Microsoft iCampus Grant, the MIT Media Lab Things that Think and Digital Life Consortia. We also thank Don Steinbrecher for his support.

All authors are from the Personal Robots Group at the MIT Media Lab, 20 Ames St. Cambridge, MA 02142 {rehtaeh, rtoscano, anjchang, wdstiehl, yiw, cynthiab}@media.mit.edu

shows a diagram of how the social touch processing system outlined in this paper will interface with the Huggable robot software.

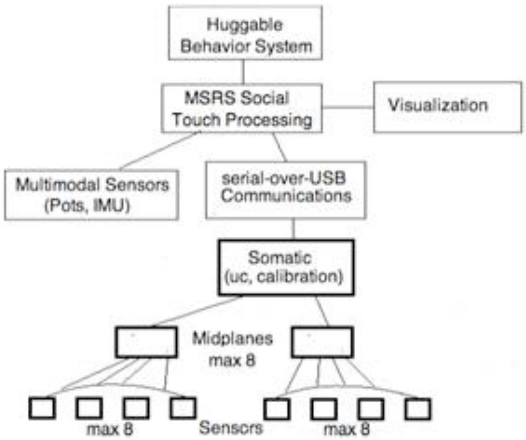

Fig. 1. Target System for Huggable: The Social Touch Processing developed on the Sensate Bear will be used in conjunction with multimodal sensors to feed information and take commands from the Huggable Behavior system. A visualization of sensor and gesture activity is also available

III. System Overview

Our algorithms mirror the structure and connotations of human touch, and we develop a simplified hardware to implement our system. The advantages of full-body social touch are differentiation of regionally significant gestures, at the expense of heavier computational load.

Real-time recognition requires tiered processing and rapid sensing. We selected capacitive sensors because they are fast, differentiate people from most objects, and sense proximity in addition to contact. The Sensate Bear uses a circuit design based upon [15], and is configured into a network of 56 modular sensors covering the entire bear. Figure 2 depicts the system components and flow.

The sensors connect through *Midplane boards* to a central *Somatic processing board* for calibration. From there, signals pass via USB to the computer where gesture recognition takes place. The microcontroller can stream signal data with 10-bit resolution. Even when treated as on-off sensors, however, our studies showed high correlation for the tested subset of Symbolic Gestures.

Once on the computer, the Sensate Bear software, created with Microsoft Robotic Developers Studio in C#, reads the associated COM port data and performs thresholding, gesture recognition, and processing, then displays active sensors and gesture classifications on a locally hosted website visualization.

During processing, we track *social touch gestures*, i.e. tactile communication or affective gestures between the human and bear. These gestures play a key role in *robot social touch*, which contrasts traditional *robot functional touch* research, e.g. object manipulation. In particular we identify *Symbolic Gestures* that have social significance across individuals and associated regional touch distributions (e.g. hug, footrub), and *touch subgestures*, which are smaller scale and are independent of location (e.g. pat, poke).

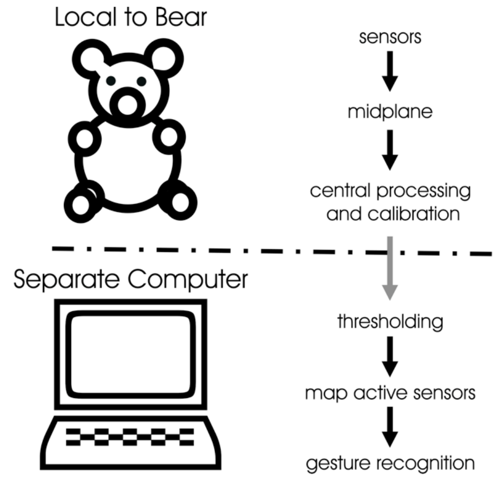

Fig. 2. System Overview: Subcomponents include sensors and electronics on the Sensate Bear, followed by on-computer processing and classification

IV. Hardware Design

A. Tiered Hardware Architecture

Within the bear, all communication and calibration takes place on a centralized *Somatic processing board*. It gathers data in a tree structure from eight *Midplane board* channels, each of which processes the signal from up to eight capacitive sensors, Figure 3. We tune our sensing circuits to detect human touch to a height of approximately 1 inch.

The micro-controller on the Somatic processing board streams data using serial over USB. It gathers information by iterating through the eight Midplanes. A more detailed description of the electronics design can be found in [15].

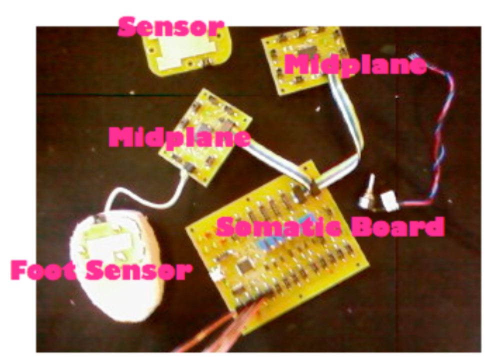

Fig. 3. Bear Electronics: Sensor board attached to foam 'foot' connects to an 8-channel Midplane, which itself plugs into one of the eight Midplane connectors on the Somatic board.

B. Physical Structure

The Sensate Bear has a rigid foam body, constructed to house the sensors under the fur of a commercial teddy bear.

Its 56 sensors are installed on the surface of a foam superstructure with body segments and shapes that mirror the physical structure of the Huggable, as shown in Figure 4. The electronics for processing are inside the head and torso.

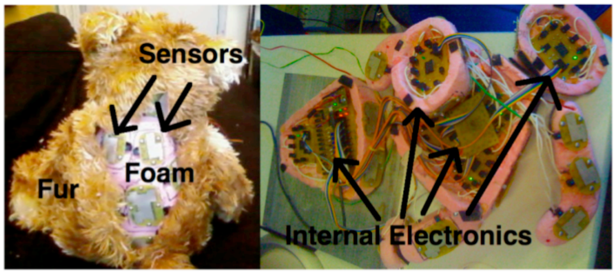

Fig. 4. The passive test-rig *Sensate Bear* has a foam superstructure that fits under a Teddy Bear's fur. Capacitive sensors span the surface of the foam and signal/communication electronics are mounted inside.

C. *Optimizing Sensor Density*

The identification of salient locations was based on a behavioral study where subjects role-played social interactions with a teddybear on videotape. The initial Huggable sensate skin design called for 1200 sensors of three different types, however, this bear simplifies the hardware design.

The previous Huggable paw had 2"x1" boards, each equipped with eight QTC pressure sensors, four temperature sensors and one electrode, which was tied to the electrodes of three other boards for region sensing.

The Sensate Bear uses capacitive sensing exclusively with two sensor sizes, 1.5"x2" and 3"x2". Sensor size and layout is designed for social touch.

D. *Capacitive Sensing for Social Touch*

Capacitive sensors are well suited to social gestures as they are fast, inexpensive and can use conductivity to differentiate human from object-based touch. The change in signal level increases with proximity, contact area and person/object conductivity. Direct touch results in the maximum change, saturating detection. To capture these edge cases (such as when a person is touching through layers of clothing), calibration is necessary.

To detect the presence of a person near the Sensate Bear we use Motorola/Freescale Semiconductor 33794 integrated circuit chips, which convert the touch's net effect on the electric field to an output voltage at a rate of five milliseconds per channel. This chip is located on the Midplane board and controls eight sensing channels. Thus, all sensors on the bear can be read in 40 milliseconds.

The sensor itself is a shielded electrode, essentially two metal plates, separated by a nonconductor. The shield faces inward, directing sensitivity outward, amplifying the signal and decreasing sensor cross-triggering and electronics interference. Because of this simple construction, capacitive sensors can be soft or hard and will ultimately be used to sense non-direct touch through the fur of the Huggable.

V. Algorithms and Software

A. *Detection Modes for Social Touch*

After analysis of our behavioral study, we targeted three classes of touch for social touch recognition; *Sensor Level touch* includes localization and activation levels, *Touch Subtypes* involve detection of base forms of local touch (e.g. poke, pat) and *Symbolic Gestures,* a novel category which posits that there are common body-wide social gestures that have typical regional distributions (e.g. hug, handshake). These processing paths are depicted in Figure 5.

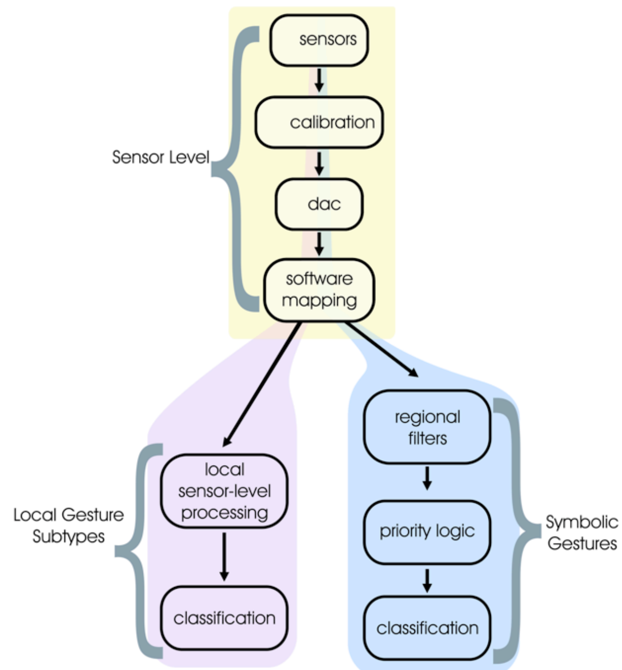

Fig. 5. Processing Pathways: The Sensor Level outputs calibrated, localized signals, which are then processed into local Touch Subtypes (time dependent) and body-wide Symbolic Gestures (not time dependent).

In sensor level touch, we read in the sensors, then process and scale their signals in calibration to use the full ground to supply voltage range. Next we convert the analog to a 10-bit digital signal and send it over USB to software, which uses a XML lookup table to map incoming sensor ID's to sensor locations.

Our detection of Touch Subtypes and Symbolic Gesture analyzes sensor activation patterns using a pre-defined touch gesture library, developed based on human studies and target behaviors of the bear. We present our first findings of what these libraries should include.

The software creates an object for each sensor that consists of: the sensor name, a buffer of sensor values reflecting the last twenty cycles, an activation state boolean that indicates whether the signal is over 30% on, and a location. Buffer size and activation thresholds are configurable.

The local touch subtype pathway uses the buffer of signal values for a single sensor to calculate features and then iteratively checks for a matching subtype in the gesture library. Only one subtype can be active on each single sensor board and the detection speed for new gestures is inversely proportional to the size of the buffer.

The Symbolic Gesture Pathway utilizes the location information of multiple sensors, passing those patterns through locational filters at each timestep, using a simple priority logic to constrain which features can be active at the same time, and updating the classification in real time.

B. *Touch Gesture Library*

We observed the gestures listed in Table I during an adult behavioral study with nine subjects (mixed gender) and a user study with eleven children (age 4-11, mixed gender). Study details and procedure are published in [14], [16]. As in Figure 5, the processing path for Symbolic Gestures is:

Locational Filters ⇒ Priority Logic ⇒ Classification

While the Touch Subtypes path is:

Fill Data Buffer ⇒ Extract Features ⇒ Classification

Noted gestures were used to verify the techniques that follow for local and body-wide gesture recognition. This list provides a cross-platform starting point for anthropomorphic creatures. Tickle appears in both categories because people tend to associate particular regions as being ticklish, but there also exists a distinct tickle subgesture that we can use to refine final classifications.

TABLE I
INITIAL GESTURE LIBRARY FOR ROBOTIC TEDDY BEARS

Symbolic Gesture	Touch Subtype
Tickle	Pet*
Footrub*	Poke*
Handshake	Tickle*
Head-pat*	Pat
Shoulder-tap	Hold*
Belly-tickle	Tap
Side-tickle*	Shake
Foot-tickle	Rub
Go-to-sleep	
Wakeup	
Feeding	
Rocking	
Dance	

*Gesture implemented

Gestures are compiled from our behavioral and user studies and extend previously identified Huggable affective touch subtypes.

C. *Sensor Level Touch*

Sensor level touch occurs mainly within the Sensate Bear electronics, which scale and condition the analog sensor signals before digital conversion. Data is passed into software as a list of sensor IDs and 10-bit amplitude levels. Each sensor ID is mapped in software to its respective on-bear location, so this format is sufficient to identify or interpolate touch locations.

Although the sensor IDs coming from the microcontroller mirror the Somatic to Midplane to sensor wiring, our software uses an XML document that delineates body regions and remaps data into human-readable labels, e.g. 'head2'.

Direct access to sensor locations enables social behaviors, such as 'Look At' in which the bear's gaze might track a touch location, as posited in [2].

D. *Touch Subtype Processing*

In this implementation, the recognition of local gestures for a single sensor is based on several seconds of data history. To classify time dependent gestures, we need reasonable but real time accuracy.

Similar classifiers have been used for the Huggable paw segment [11] and in pressure-based skins [10]. Features in the first (paw segment) include: direction of motion, average sensor value, change in sensor value, number of sensors active; and in the second (pressure sensors): absolute values, spatial distributions and temporal difference.

In keeping with our observation-based design, we selected and added to these heuristics based on three subjects' demonstration of tickle, poke, and pet. Figure 6 shows the raw oscilloscope traces of these subtypes.

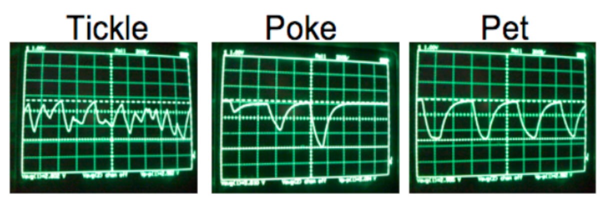

Fig. 6. Typical Touch Subtypes: Raw oscilloscope captures of subject demonstrating tickle, poke and pet on a single 3x2" sensor board. Testing physical touch, not proximity. Duration is 5 seconds.

For each subtype (Table II), we evaluated peak amplitude, base frequency, frequency spectrum and duration, the features that best separated our dataset.

TABLE II
TRIAL SUBGESTURE FEATURE RESULTS

AVERAGE	Tickle	Poke	Pet	Hold	No Touch
Peak Amplitude	60%	>30%	100%	100%	0%
Base Frequency	5-10Hz	0-1Hz	0.5-2Hz	0 Hz	0 Hz
Freq Spectrum	high noise	blip	low noise	no noise	no noise
Duration	3-20 sec	1 sec	>4 sec	>4 sec	n/a

STD DEV	Tickle	Poke	Pet	Hold	No Touch
Peak Amplitude	10%	30%	10%	0%	0%
Base Frequency	5 Hz	0.5 Hz	0.75 Hz	0 Hz	0 Hz
Freq Spectrum	large	n/a	small	0	0
Duration	8 sec	0.2 sec	15 sec	20 sec	n/a

Averaged data and standard deviations based on ten iterations with each of three users.

Signal amplitudes are highest in pet and hold. In taking the

Fourier transform of those signals, there is variation in the dominant base frequency and overall distribution of frequencies (noise level).

As part of the touch subtype processing, a variable in the sensor class stores the last several cycles of values for its activation level as an array. We calculate feature values from that history. The peak amplitude is the maximum value in the set, base frequency can be calculated from lowest frequency value in the Fourier Transform, noise level reflects frequency spread, and duration is incremented for each cycle. Pseudocode for this process follows:

```
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;
     }
}
```

E. Symbolic Gesture Processing

Symbolic Gestures track regionally-significant touch interactions in real time. Touch location on a body is highly tied to social intention, as found in [14], particularly given the anthropomorphic profile of this robot. Thus, our algorithms use a *locational filter* for each Symbolic Gesture (see Figure 7) gleaned from human pattern recognition of the adult behavioral study.

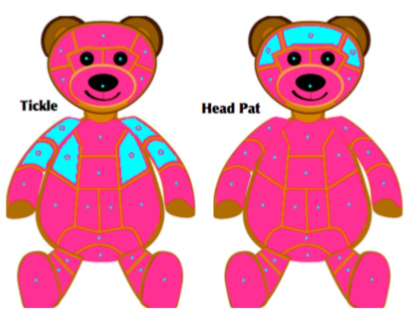

Fig. 7. Locational filter examples for Symbolic Touch: The programmed side-tickle and head pat filters are visualized here.

At each cycle, the software tests for any matching activation patterns, displaying active gestures on screen. The processing is probabilistic, a minimum number of sensors within the locational filter must be active for at least two seconds, an increment informally chosen to parallel human recognition time.

Next, the algorithm enters its *priority logic*. It is possible to have multiple gestures, but we must capture the cases were classifications conflict. For example, the tickle distribution involves many of the same sensors as hug. However, hug involves various other unique sensors. Thus, if hug is active, tickle is unlikely to be happening, so hug supersedes Tickle. Priority logic must be based on human behavior and recalculated when adding new gestures.

Labeled interactions for the initial implementation are headpat, hug, tickle, and footrub. These classifications were approximations of user behavior from the behavioral study and represented the most used expressive gestures therein. The logic for each is as follows:

Tickle: (not Hug) && (active >= two of four sensors)
Headpat: (active >=one of the three sensors)
Hug: (both sides)&& (active >= four of ten sensors)
Footrub: (active >= one of two sensors)

If needed, the robot can also evaluate the subtypes present within any active distribution.

At a higher level, the robot behavior system will eventually associate affective and communication content with Touch Subtypes (poke always gets attention) and Symbolic Gestures (hug has a positive reassuring effect).

VI. Results

A. Timing Results

Figure 8 depicts the timing delays during data flow from sensors over serial then in the software classification paths. The total per program cycle time is about 42 milliseconds including communications delay – thus locational filters can be processed ~20 times a second and subtypes are assessed about every second. Based on the observed gesture lengths in the study (see Table II), that is similar to human recognition time.

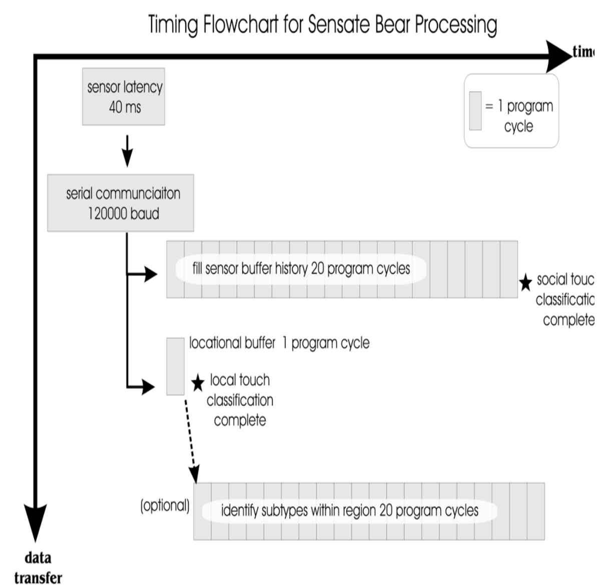

Fig. 8. Timing Flowchart for Processing: Each program cycle has a sensor latency, communication delay, and runtime which limits to minimum time before categorization complete.

B. Touch Subtype Results

Our experiments informed the touch subtype classification features, using the values from Table II. We used observed data to craft an algorithm and verify it piecewise.

Because the capacitive sensing mux requires time to acquire signal after switching channels, all sensors are queried in turn with a 5-millisecond delay between each one. Thus the true data rate is (58 sensors) x (.005 seconds) = 0.29 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.

Classification of subtypes has been done in prior work [11]. What is novel is that we implement this for the first time exclusively with capacitive sensing.

C. Symbolic Gesture Results

We evaluated the most common Symbolic Gestures to verify if the probabilistic locational filters correctly reflect a user's touch during the user study with children. By instructing subjects to demonstrate these gestures in the context of a social interaction, we acquired the results shown in Table III.

TABLE III
FIRST TEST OF REGIONAL FILTERS

	Activated on first try	**Activated with explanation**	**Regional Accuracy**
Headpat	100%	100%	100%
Tickle	20%	60%	20%
Hug	40%	80%	80%
Footrub	100%	100%	100%

Subjects instructed to demonstrate labeled gestures. High activation rate is due to the lack of crossover between locations of touch in the highlighted gestures.

We observed, in confirmation with [14] that particular regions of the bear tended to have associated social content, regardless of subgesture. Thus, anthropomorphic profiling may already provide a higher than chance likelihood that particular social gestures are present.

VII. CONCLUSIONS

This paper is about developing a real time system for social and affective touch. Key contributions are, first, that our algorithms are based on real humans interacting with a plush and then sensate bear. Second, our research with the Sensate Bear test rig will ultimately be incorporated into the Huggable personal robot system. Third, we motivate research into regionally symbolic touch research, making use of the full-body sensing and anthropomorphic nature of the bear.

We have presented our in-development approach and results for realtime classification. Table II demonstrates feature data distributions that distinguish affective touch sub-types. Table III describes a study to demonstrate the regional nature of social touch.

Now that we have outlined an approach to real time social and affective touch, the next step is to combine the local and symbolic touch recognition pathways. We will also use training data sets to further evaluate our algorithms and flesh out our gesture library.

The preliminary timing results indicate that our system will be capable of real-time touch, demonstrating a tiered approach to touch recognition.

Our taxonomy of touch includes *Symbolic Gestures* that have associated locational distributions, *Touch Subtypes* that are locationally independent, and *Sensor Level touch*. Each of those can be accessed in software.

Thus, we have also created the software base for a Robotic Touch API, showing that the Huggable behavior system will be able to query currently active sensors and values, gestures and subgestures.

REFERENCES

[1] Field, T. 2001. Touch. MIT Press.
[2] Sherwyn, D., Lynn, M. & Le, J. 1998. Reach Out and Touch Your Customers. Cornell Hotel & Restaurant Administration Quarterly, June 1.
[3] Parshall, D. 2003. Research and reflection: animal-assisted therapy in mental health settings. Counseling and Values. Oct 3.
[4] Wada, K., Shibata, T., et al. 2005. Robot Assisted Activity at a Health Service Facility for the Aged. IEEE Workshop on Advanced Robotics and its Social Impacts.
[5] 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.
[6] Kanda, T., T. Hirano, et al. 2004. Interactive Robots as Social Partners and Peer Tutors for Children. Human Computer Interaction 19(1/2): 61-84.
[7] Thayer, S. 1982. Social Touch, in Schiff, William, and Emerson Foulke Tactual perception: a sourcebook. Cambridge University Press, pp. 263-304
[8] T. Tajika, et al, Automatic Categorization of Haptic Interactions - What are the Typical Haptic Interactions Between a Human and a Robot?, Humanoids 2006.
[9] F. Naya, J. Yamato, K. Shinozawa, Recognizing Human Touching Behaviors using a Haptic Interface for a Pet-robot, SMC'99.
[10] M. Hayashi, et al., Development of Functional Whole-Body Flesh with Distributed Three-axis Force Sensors to Enable Close Interaction by Humanoids, IROS 2007.
[11] W. D. Stiehl and C. Breazeal, "Affective Touch for Robotic Companions," in First International Conference on Affective Computing and Intelligent Interaction, Beijing, China, 2005.
[12] Stiehl, W.D. et al. 2005. Design of a Therapeutic Robotic Companion for Relational, Affective Touch. IEEE Int'l Workshop on Robot and Human Interactive Communication (RO-MAN 2005), pp. 408-415.
[13] Toscano, R. L, Lee, J.K., Stiehl,W.D.; Breazeal, C. 2008. The design of a semi-autonomous robot avatar for family communication and education. Robot and Human Interactive Communication, pp.166-173.
[14] Knight, H. 2008. An Architecture for Sensate Robots: Real Time Social-Gesture Recognition using a Full Body Array of Touch Sensors. Master's Thesis, Dept. of EECS. MIT.
[15] Stiehl, W.D., Breazeal, C. 2006. A Sensitive Skin for Robotic Companions featuring Temperature, Force, and Electric Field Sensors. Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems
[16] Knight, H., Chang, A, Breazeal C. et al. 2009. Robot Design Rubrics for Social Gesture Categorization and User Studies with Children. HRI Workshop on Acceptance of Robots in Society.