Affective Touch for Robotic Companions

Walter Dan Stiehl and Cynthia Breazeal

Robotic Life Group, MIT Media Lab, Cambridge, MA 02139, USA
wdstiehl@mit.edu, cynthiab@media.mit.edu

Abstract. As robotic platforms are designed for human robot interaction applications, a full body sense of touch, or "sensitive skin," becomes important. The Huggable is a new type of therapeutic robotic companion based upon relational touch interactions. The initial use of neural networks to classify the affective content of touch is described.

1 Introduction

The realm of touch in robotic systems has largely been limited to manipulation. In these platforms, such as the NASA/DARPA Robonaut Hand [1], tactile sensors are only placed on the surface of the robotic hand or gripper, and the remainder of the robot is largely not sensed. Clearly, there are many other applications for tactile sensing in robotics.

Lumelsky, Shur, and Wagner originally proposed the idea of a "sensitive skin" in 2001 [2]. A "sensitive skin" features a large number of sensors of different modalities, such as temperature, pressure, etc., which cover the entire surface of the robot. Biological skin also features a large number of sensors of different modalities for the purpose of encoding the external world through touch [3]. Currently, there exist only a handful of robots which feature a full-body "sensitive skin," such as [4]. In many of these applications, the "skin" is used primarily to protect the robot from damaging itself or a human user. There are a number of other applications for such "skins" which have yet to be addressed.

In this paper, we present the Huggable, a new type of therapeutic robotic companion based upon relational touch interactions. Unlike other robotic companions, such as Sony's AIBO [5], the Huggable features a *full body*, multi-modal, "sensitive skin" for the purpose of sensing the affective content of touch in human-robot interactions. Unlike previous work, which has used a neural approach for object recognition through touch [6], the focus of this approach is to detect the affective content of the social interaction – petting, tickling, etc. A brief description of the design of the Huggable is first presented to provide context for the neural network approach used to classify the affective content of social tactile interactions, the focus of this paper. A more detailed discussion of the design can be found in [7].

2 The Huggable: A Therapeutic Robotic Companion for Relational, Affective Touch

In the full implementation of the Huggable, shown in Fig. 1, it features a full-body sensitive skin, silent back-drivable voice coil actuators [8], an inertial measurement

J. Tao, T. Tan, and R.W. Picard (Eds.): ACII 2005, LNCS 3784, pp. 747 – 754, 2005.

unit [9], and an embedded PC with wireless communication capabilities for behaviors, patient monitoring, and data collection. Vision and auditory processing may be added as well to allow for multi-modal interactions.

Fig. 1. The Huggable – Currently in Development

The goal of the Huggable project is to deploy these robotic companions to hospitals and nursing homes for robot therapy applications with a wide range of users from the elderly to small children. As such, a series of design decisions were made to develop a robot which invites users of all ages to interact with it. First, the overall weight of the Huggable must be kept below 5 lbs to allow it to be picked up and held in one's arms. The Huggable features a strong, light-weight, ribbed mechanical understructure for this purpose. Second, the Huggable must have a soft feel that is pleasant to touch. Unlike many robotic platforms which have a hard exterior, the Huggable features a soft silicone synthetic skin under the soft furry exterior of Fig. 1.

Third, the Huggable must convey the "illusion of life" through life-like motion. Quiet, smooth, back drivable voice-coil actuators, as opposed to the traditional geared DC motors, are used for all the degrees of freedom of the Huggable. In the current design, the bear features a 3-DOF neck, 2-DOF eyebrow mechanism, 1-DOF ear mechanism, and a 2-DOF "Hug" shoulder mechanism. The neck and shoulders allow for active touch behaviors such as orienting towards touch, nuzzling, and hugging. The ears and eyebrows are used to express the internal state of the robot.

Finally, the Huggable must be able to detect the affective content of the social interaction the user is having with the robot, and respond appropriately. To accomplish this design constraint, the Huggable features a full body "sensitive skin" based upon the ideas of [2], as well as the organization of the somatosensory system in humans and animals [3]. The "sensitive skin" of the Huggable features four modalities of somatic information – pain, temperature, touch, and kinesthetic information. For touch, electric field sensors are used to measure proximity and Quantum Tunneling Composite (QTC) sensors are used to measure force. Temperature is sensed using thermistors. Kinesthetic information is measured by potentiometers. Finally, pain is treated as an intense sensor signal of any one of these stimuli. Due to space constraints, a further discussion of these sensors and the processing used appears in [7].

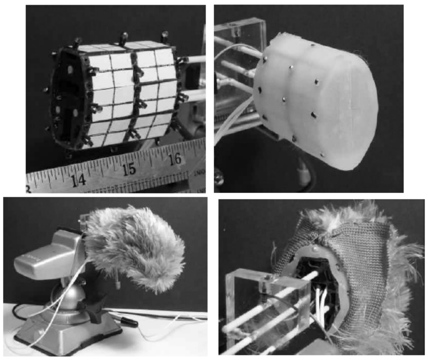

Fig. 2. The Finished Arm Test Section. The upper left shows the sensor layout. Each white rectangle is one QTC sensor. The silver sensors above the surface of the QTC sensors are the thermistors. The electric field sensing electrodes, not shown, are located on the bottom plane of each sensor circuit board. The upper right shows the synthetic silicone skin placed onto the assembly. The holes cut out in the skin allow for the thermistors to pass through the skin for better temperature sensing. The lower left is the finished assembly with the fur arm sleeve attached. The lower right shows the layered structure. All dimensions shown are in inches.

Currently, a single arm section, shown in Fig. 2, has been created. It features a total of 64 QTC sensors, 24 temperature sensors, and 2 electric field sensing electrodes. The finished Huggable will have over 1000 QTC sensors, 400 temperature sensors, and 45 electric field sensing electrodes. This high spatial resolution is necessary to detect the social and affective content of touch.

3 Classifying the Affective Content of Touch

There are many types of touch interactions that convey affective or social content. Handshakes are a form of greeting. Poking is used to get someone's attention. Petting is a calming gesture for both the animal and the person doing the petting. Slaps are a form of punishment. Many more examples exist. The Huggable is designed for therapy applications for people who have limited or no access to companion animals. Thus, the Huggable should be able to classify the numerous types of tactile interactions that people have with companion animals, so as to respond appropriately. This section describes an early experiment conducted with the arm section of Fig. 2 to classify the affective content of touch.

3.1 Experimental Design

The arm section of Fig. 2 was clamped to the top surface of a lab bench. Data from the top and left region of the arm section (half of the arm) was recorded as a baud rate of 57600. This arm section was selected as it corresponded to one electric field electrode. A total of 200 data sets were created from 16 different affective touch interaction types. These interactions were organized into 9 classes and 6 response types as

Table 1. The Data Sets Used for the Classification of Affective Touch

Type	# of Data Sets	Class	Response
Tickle: Softly, Fingers Only	10	Tickle	Tease Pleasant
Tickle: Hard, Fingers Only	10	Tickle	Tease Painful
Poking: Softly	20	Poking	Tease Pleasant
Poking: Hard	20	Poking	Tease Painful
Scratching: One Finger Softly	20	Scratching	Touch Pleasant
Scratching: One Finger Hard	20	Scratching	Touch Painful
Slapping: Fingers Only Softly	10	Slapping	Punishment Light
Slapping: Fingers and Palm Softly	10	Slapping	Punishment Light
Slapping: Fingers Only Hard	10	Slapping	Punishment Painful
Petting: Softly	10	Petting	Touch Pleasant
Petting: Hard	10	Petting	Touch Painful
Patting: Softly	10	Patting	Touch Pleasant
Patting: Hard	10	Patting	Touch Painful
Rubbing	10	Rubbing	Touch Pleasant
Squeezing	10	Squeeze	Touch Painful
Contact	10	Contact	Touch Pleasant

shown in Table 1. The response type is how the Huggable interprets what behavior to perform. For example, a pleasant touch should signify a happy reaction while strong punishment should result in a pain response. In this preliminary study all data sets were created through interaction with the arm section by a single person. Future work will expand upon this with multiple users.

Due to time constraints, the sensor values were not calibrated. Equation 1 was used to normalize each sensor reading to a baseline for each data set of Table 1:

$$NormalizedSensorValue = \frac{SensorValue - baseline}{SensorMaxValue - baseline} \tag{1}$$

The reading of sensor values was not optimized in time. The electric field sensors required a much longer time period to read a single value, 4 ms, compared to the temperature and force sensors, 50 µs. Thus in the current implementation the electric field sensors created a bottleneck as each sensor was read one after another. In future, it will be possible to read from all the temperature and force sensors in a body region while the electric field sensor is being read, thus eliminating this problem.

3.2 Feature Extraction

A set of features are calculated from each data set at each time step. The normalized sensor value of Equation 1 is filtered and negative values are set to zero. The number of active sensor values within a receptive field is calculated and normalized by the number of sensors within that receptive field. For example, if the receptive field consists of the 8 QTC sensors on a single sensor circuit board and all sensors are active then the value of the active sensor count is 1.0. If only 4 of the 8 sensors are active in one receptive field, this value would be 0.5.

The normalized sensor value sum is calculated by dividing the sum of the sensor values in each receptive field by the number of sensors in that field. The normalized sensor value sum is 1.0 if all the sensors in a receptive field are at their maximum value. Thus, this feature calculates the average sensor value for a single receptive field. In contrast, the normalized average sensor value sum is calculated by dividing the sum of the sensor values in each receptive field by the number of active sensors in that field. This feature calculates the average across all active sensors in a field, not the total number. The normalized average sensor value sum is 1.0 if all the active sensors in a receptive field are at their maximum value and the inactive sensors were 0.

The change in sensor value sum between each time step is calculated by subtracting the previous normalized sensor value sum from the current normalized sensor value sum. Two additional binary features are used to indicate the direction of the change. The sensor value increasing feature is true (1) if the sign of the change in normalized sensor value sum is positive. The sensor value decreasing feature is true (1) if the sign is negative.

A set of binary features is calculated to determine the degree of activity within a receptive field. One feature is a 1 if the current sensor was active at the last time step and still is. Another feature is 1 if the sensor was not active at the last time step and still is inactive. Finally, one other feature is 1 only if there was no change in sensor value, i.e., the sensor was inactive and still is, or was active and still is. The number of time steps since there was a change in activity for a sensor is calculated. This value is normalized at 1 if it has been 100 samples or more since there was a change in activity (i.e., changing from active to inactive or vice versa). This feature can encode the periodicity of a signal.

The centroid location and direction of motion within a receptive field are calculated as described in [10]. The direction of motion is then divided into eight 45-degree regions – up, upper left, left, lower left, down, lower right, right, and upper right. A corresponding binary feature for each of the eight directions is 1 if the direction of motion falls within that range and 0 otherwise. Changes in the direction of motion are indicated by another binary feature. Finally, the number of time steps since the last direction change is calculated.

The features described in the section can be calculated for any receptive field. A receptive field could be the entire arm, or simple one sensor processing board within the arm. The receptive field could also consist of all the types of sensors – QTC, temperature, and electric field; or could be one specific sensor type. In this paper features were calculated for each sensor type, with the receptive field being a single sensor circuit board consisting of 8 QTC sensors, 3 temperature, and a shared electric field sensor.

3.3 Neural Network Classification

The 200 trials of Table 1 were randomly assigned into a training and validation data set of approximately equal size. One of the data sets of Table 1 was corrupt, thus there were only 199 data sets used. MATLAB was used for feature extraction as well as training and testing of the neural network offline. A neural network approach was chosen for classification in this preliminary study.

A three layer neural network was trained with 100 inner layer nodes. The number of output nodes was 16 for the type neural network classifier, 9 for the class neural network classifier, and 6 for the response neural network classifier. The "logsig" transfer function was used with the "trainrp" transfer function for all three neural networks. The learning rate was 0.001. The maximum number of epochs was 1000. The error tolerance was 1e-3. The output of each neural network was rounded, so that all probabilities greater than 0.5 were recorded as being classified and those less than

0.5 were not classified. Due to time constraints, only one neural network was trained for each of the type, class, and response classifiers. Table 2 and 3 show the results of the neural network classification for class and response classification. The classification by type is not shown due to space constraints.

There were a large number of "no contact" situations within the data sets. Thus, the usual measurement of accuracy does not reveal much information as the number of true negatives, due to "no contact" or other affective touch situations, dominate the number of true positives in the data set, yielding very high accuracies. Specificity, sensitivity, positive predictive value (ppv), and negative predictive value (npv) are used instead [11]. Specificity, the probability that an interaction will be classified as not occurring when that interaction did occur, is given by Equation 2. Sensitivity, the probability that an interaction will be classified as having occurred when it did occur, is shown in Equation 3. The positive predictive value, Equation 4, and negative predictive value, Equation 5, is the probability that the interaction did or did not occur.

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive} \tag{2}$$

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative} \tag{3}$$

$$ppv = \frac{TruePositive}{TruePositive + FalsePositive} \tag{4}$$

$$npv = \frac{TrueNegative}{TrueNegative + FalseNegative} \tag{5}$$

The classification by class yielded much better results than the classification by type. This was expected due to the fact that the distinction between "soft" and "hard" versions of the same stimuli may not have differed sharply as force was not controlled during data collection. Thus it may have been that the same stimuli could have been labeled as "soft" and "hard." As shown in Table 2, for all cases, except for slap (not shown), the positive predictive value and sensitivity are greater than chance. Slap was not classified well by the neural network indicating an error in the design of the feature detector, or there may have been very few occurrences of labeled slaps in the training set. A slap happens very quickly and only a single time, compared to the other interactions such as scratching or rubbing. In the current implementation, there is a 4 ms delay per cycle as the electric field sensor is read. As such, the amount of slap data on which the neural network could be trained is much less than the other types of interactions.

As shown in Table 3, the performance of touch classification showed the best results. The classification of punishment, light or strong, is not shown in the table due to its very poor classification. As discussed previously, the classification of the slap class yielded very poor results. The punishment response type consists primarily of slaps as shown in Table 1. Thus, it makes sense that the punishment response would be classified very poorly as well. An improvement in the classification of slaps should yield a better classification of the punishment response.

The grouping of classes into response types is done subjectively. Thus there may be a better grouping which yields better results. Another interesting pattern emerges from looking at the grouping of the types into responses shown in Table 14-1. The touch pleasant has five types. The touch painful has four types. The tease pleasant, tease painful, and punishment light each have two types. The punishment strong only has one type. There is a strong relationship between the order of positive predictive value and sensitivity to the number of types within one response. Thus it appears that the amount of data within the training set is having an effect on the classification result. In the future the responses should be better balanced in terms of quantity of data for each response in the training set.

Table 2. The Results of the Neural Network Classification for Class. PPV = positive predictive value, NPV = negative predictive value.

Class	PPV	NPV	Sensitivity	Specificity	Chance
Tickle	0.67	0.94	0.57	0.96	0.11
Poke	0.41	0.95	0.29	0.97	0.11
Scratch	0.67	0.94	0.65	0.94	0.11
Pet	0.58	0.97	0.26	0.99	0.11
Pat	0.23	0.99	0.20	0.99	0.11
Rub	0.72	0.98	0.73	0.98	0.11
Squeeze	0.84	0.97	0.73	0.98	0.11
Contact	0.81	0.98	0.84	0.97	0.11

Table 3. The Results of the Neural Network Classification for Response Type. PPV = positive predictive value, NPV = negative predictive value.

Response	PPV	NPV	Sensitivity	Specificity	Chance
Tease Pleasant	0.40	0.93	0.25	0.97	0.17
Tease Painful	0.66	0.94	0.53	0.96	0.17
Touch Pleasant	0.77	0.90	0.74	0.90	0.17
Touch Painful	0.70	0.91	0.70	0.91	0.17

4 Conclusions and Future Work

The results shown for the classification of affective touch show promise for the creation of such systems for a wide variety of robotic platforms. The initial use of a neural network approach shows that such social affective classes of touch can be separated from one another. The next steps will be to continue to develop the Huggable with the goal of the full body implementation of the "sensitive skin" and to develop other real-time classification methods for the affective content of touch. As the robot is able to process how it is being held and interacted with, it can select the appropriate behaviors and motion to provide a very enriching experience for those who do not have the access to companion animal therapy.

Acknowledgments

The authors wish to thank Professor Chris Moore of MIT Brain and Cognitive Sciences for his help in the understanding of the somatosensory system. Professor Joe

Paradiso of the MIT Media Lab provided help with the implementation of the electric field sensor. Professor Rosalind Picard of the MIT Media Lab provided an introduction to pattern classification techniques through her class. The authors also thank the other members of the Robotic Life Group, specifically Jeff Lieberman, Louis Basel, Levi Lalla, and Mike Wolf who also have contributed to the design and construction of the Huggable. This work is partially supported by a Microsoft iCampus grant, the MIT Media Lab Things that Think and Digital Life Consortia, and the NSF Center for Bits and Atoms Contract No.CCR-0122419.

References

1. Martin, T.B., et al. Tactile Gloves for Autonomous Grasping with the NASA/DARPA Robonaut. in International Conference on Robotics and Automation (ICRA'04). 2004. New Orleans, LA.
2. Lumelsky, V.J., M.S. Shur, and S. Wagner, Sensitive Skin. IEEE Sensors Journal, 2001. 1(1): p. 41-51.
3. Kandel, E.R., J.H. Schwartz, and T.M. Jessell, Principles of Neuroscience, 4th Edition. 4th ed. 2000, New York: McGraw-Hill Health Professions Division. 1414.
4. Iwata, H. and S. Sugano. Whole-body Covering Tactile Interface for Human Robot Coordination. in International Conference on Robotics and Automation. 2002. Washington, DC: IEEE.
5. Sony Product Literature, AIBO ERS-7M2 website: http://www.sonystyle.com/is-bin/INTERSHOP.enfinity/eCS/Store/en/-/USD/LC_BrowseCatalog-Start;sid=5PwkGv1dhiQkIb8nCd8uEbJRgVs-2uQV4R0=?CategoryName=lc_AIBO.
6. Heidemann, G. and M. Schopfer. Dynamic Tactile Sensing for Object Identification. in International Conference on Robotics and Automation (ICRA'04). 2004. New Orleans, LA: IEEE.
7. Stiehl, W.D., et al. Design of a Therapeutic Robotic Companion for Relational, Affective Touch. in IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2005) (Currently Under Review). 2005. Nashville, TN.
8. McBean, J. and C. Breazeal. Voice Coil Actuators for Human-Robot Interaction. in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS04). 2004. Sendai, Japan.
9. Morris, S.J., A Shoe-Integrated Sensor System for Wireless Gait Analysis and Real-Time Therapeutic Feedback, in Health Sciences and Technology Sc.D. Thesis. 2004, MIT: Cambridge.
10. Stiehl, W.D. and C. Breazeal. Applying a "Somatic Alphabet" Approach to Inferring Orientation, Motion, and Direction in Clusters of Force Sensing Resistors. in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS04). 2004.
11. rapid-diagnostics.org, Accuracy of Diagnostics Tests: http://www.rapid-diagnostics.org/accuracy.htm.