Gesture Recognition in the Haptic Creature

Jonathan Chang, Karon MacLean, and Steve Yohanan

Department of Computer Science, University of British Columbia
2366 Main Mall, Vancouver, B.C., V6N 2K6, Canada
jchang86@interchange.ubc.ca, {maclean,yohanan}@cs.ubc.ca

Abstract. Touch is an important but poorly studied aspect of emotional communication. With the *Haptic Creature* we are investigating fundamentals of affective touch. This small robot senses the world solely by being touched via a force-sensing resistor network, and communicates its internal state via purring, stiffening its ears and modulating its breathing and pulse. We describe the Creature's first-generation gesture recognition engine, analyze its results, and specify its next iteration. In the region of highest sensor density, four gestures were differentiated with an average of 77% accuracy. Error patterns suggest that sensor deficiency rather than algorithm pose current performance limits.

Keywords: Affect, touch sensing, human-robot interaction, gesture recognition.

1 Introduction

Affective touch communicates or evokes emotion. In the Haptic Creature project we are investigating affective touch in social human-robot interactions, to identify its physical traits for eventual applications such as therapy and companionship. Our immediate goals are the display and recognition of affective touch by human and machine, as well as the interactive touch dynamics that can develop between them [1].

We are leveraging research in human-animal interaction with a robotic creature that mimics a small animal sitting on a person's lap (Fig. 1). The Haptic Creature interacts entirely through touch by breathing, purring and stiffening its ears in response to the user's touch. We use an animal platform to avoid confounding factors in human-human social touching such as gender, social status and culture. With studies now in progress we are exploring essential traits of this form of touch, and mechatronics and computation needed to support them. What touch gestures do humans most naturally use to express specific emotions? What is required to *elicit* (form factor, surface textures, movements) and recognize them (sensing, modeling)?

This paper describes our first-generation Gesture Recognition Engine (GRE), an essential part of a platform that will help us answer these questions.

1.1 Background

Social touch. Our interest in affective touch is informed by studies of human-to-human social touch. Hertenstein et al examined touch-only communication of specified emotions between strangers, and identified the specific tactile behaviors used [2].

A.M.L. Kappers et al. (Eds.): EuroHaptics 2010, Part I, LNCS 6191, pp. 385–391, 2010.

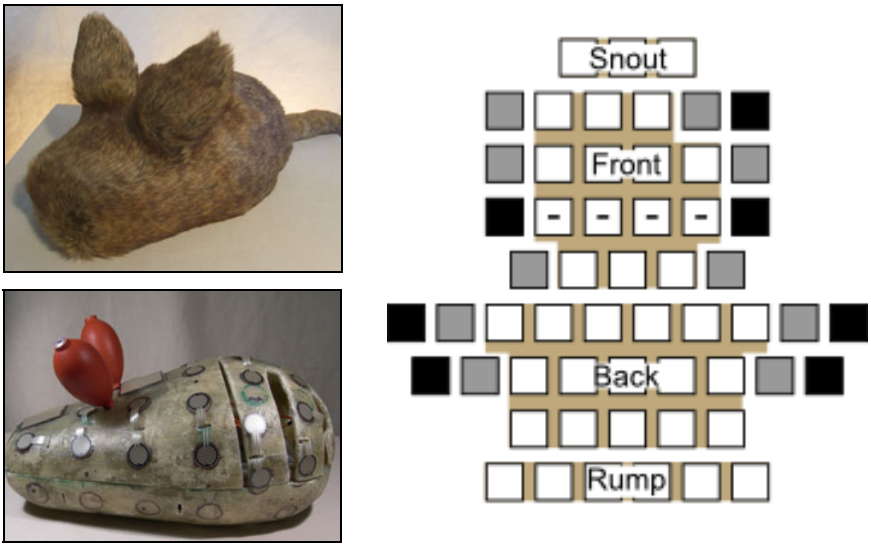

Fig. 1. The Haptic Creature and its sensors. CCW from top left: (a) with furry skin; (b) fiberglass shell and touch sensors; (c) touch sensor mapping to region, flattened; black squares indicate underbelly, and gray indicate lower sides.

Participants transmitted emotional meanings (anger, fear, disgust, love, gratitude, and sympathy) with accuracy rates of 48-83%/chance=25%), comparable to performance observed for facial displays and vocal communication.

To illuminate how touch is used for communication in real-life settings, Jones and Yarbrough asked participants to record details of touch-based social interactions and context in their daily lives; then segmented these touches into 12 distinct categories of meaning such as inclusion, sexual interest, compliance or playful aggression. For example, attention-getting usually involved "spot touches" by the initiator's hand on a non-vulnerable body part of the recipient [3]. While many questions remain and results such as these are almost certainly culturally specific, they confirm the importance of touch as a communicative medium and reveal key social mechanisms.

Human Robot Interaction and Touch. Current trends in human-robot interaction (HRI) are represented in Dautenhahn's survey of social and interactive skills needed for robots to be socially intelligent companions or therapeutic agents [4]. Touchability has been a part of a several HRI platforms, beginning with Paro [5].

Using a teddy bear with 56 capacitive touch sensors, Knight et al distinguish "touch subgestures" (low level touches such as *pet*, *stroke*, *pat* and *hold*) and "symbolic gestures" (location-dependent, with contextual social meaning such as, for this teddy bear model, *feeding*, *rocking*, *hug*, *head-pat* etc) [6]. Touch subtypes are inferred with classifiers (e.g. average sensor value, number of sensors active, spatial distributions), and symbolic gestures with priority logic relying on location filters. Training data files collected on 4 subgestures from one user produced distinctive variable profiles, but results are not reported for subgesture recognition rate, nor real-time interactive configuration discussed.

Our Approach. The Haptic Creature platform is designed for purely haptic bi-directional interaction, and takes a deliberately non-representational physical form [1]. Our GRE uses a probabilistic Markovian model with architectural similarities to [6]; the Creature has been assessed in an operational setup for recognition success of gestures analogous to [6]'s "subgesture" class. The GRE is a component of a physically interactive system, mandating eventual realtime performance and a modular, scalable structure able to concurrently evaluate multiple candidate gestures of different durations and urgencies. These recognized gestures guide the Creature's emotion model and physical behavior, which is being independently developed.

1.2 Mechatronic Description

The Haptic Creature's mechatronics are designed to produce organic, coordinated behavior with a minimalist zoomorphism that suggests animal traits while avoiding replication of a single real animal. The robot is clothed with faux fur on a fiberglass shell which supports 56 Interlink™force sensing resistors (Fig. 1(a-b)).

Fig. 1(c) shows the sensor mapping to the touch regions of interest in the GRE, with 47 round (0.5") and 9 square (1.5"); because FSR sensitivity drops in bending, small sensors were used in higher curvature regions. FSRs are spaced on ~2" centers (average), front to back and left to right, but with variations primarily dictated by curvature. Each ear bulb has two sensors, on the front and outer side.

The Creature's display features are described in [7]. The stiffness of two inflation bulb "ears" are individually controlled by out-take valves. The Creature purrs through an offset weight attached to a motor shaft spinning under modulated motor control, and simulates breathing via a servo that displaces the articulated breastplate.

1.3 System Goals and Proof of Concept

The GRE must eventually identify and differentiate 10-20 affectively significant gestures (not emotions), varying in aspects such as duration, localization, and temporal/spatial force profile. It must do this in realtime: "sharp" gestures (e.g. poke or slap) within a few milliseconds, and gentle, slow touches over a longer period.

Perfect GRE accuracy is unlikely to be feasible given the imprecision found even with people touching people. We have provisionally estimated a useable accuracy target of 75% / chance=25% (to be confirmed in later experimental stages), and have prioritized understanding affective gesture classes and correctly identifying them. Accuracy will always be best when the GRE is individually tuned, just as we best understand touches from familiar living touch-partners, such as a family member or a pet. Personal customization of the GRE is impractical for our short-term purposes, although it may be helpful for applications where the Creature will work in a dedicated partnership.

The GRE version described is a proof of concept, which we assess for feasibility of general approach. To that end, we consider (a) best-case results, which show what is possible and help distinguish mechatronic and algorithmic causes; and (b) analyze lesser results and (c) discuss the algorithm's scalability.

2 The Creature Gesture Recognition Engine

The GRE is a software system for recognizing atomic user actions through touch sensor activations. It extracts key features from the data, after normalization and averaging to deduce user intention. As a part of the overall Haptic Creature system, the mechanisms that exist within the GRE represent the Creature's internal disposition towards various kinds of touch. Just as a sensitive person may mistake your pat for a slap, the GRE's internal values may be tuned to be more or less sensitive.

The GRE passes its estimate of the most likely recent gesture to the Haptic Creature's "emoter", which determines a *response* to the identified gesture. Whereas the GRE is tunable in classification sensitivity, the emoter encodes higher-level strategies, e.g. mirroring an emotion versus provoking an emotion change in the user.

2.1 GRE Description

The Java-based GRE version described here is built with Markovian principles in mind, but in its current stage it more resembles a deterministic decision tree. The GRE can map inputs of a certain domain to outputs of another discrete domain with error probability calculations [8]. It maps the sensor array data to various gestures that caused the patterns in the sensor array, and makes error inferences based on deviations from the set gesture definitions. In Markovian terminology, the gestures represent the "state of the world", the sensory array is the "observation", and the GRE is the "agent".

The sensor array (56 10-bit values) generates an extremely large raw data stream, currently sampled at 15 Hz following Nyquist criteria applied to anticipated traversal and tapping rates (much higher sample rates are possible), and with it the danger of temporally missing quick touches altogether. To handle this stream, we created "features" [8] which extract key data properties. These can be thought of as key statistics (peak activation, average activation, points of contact, etc.) that summarize the raw data to ease processing.

Features employed in the current evaluation (Table 1) were chosen pragmatically, based on greatest accuracy for the gestures targeted below. Additional features still in development are *gesture vector* (direction and intensity of the activated sensors); and *centroid movement* (movement of an activation patch's centroid over time). Less helpful to date are *peak activation area, peak movement*, *median and quartile activation and area*, and *centroid activation* (the level of activation at the centroid).

Table 1. Feature values used and their levels.

Feature	Description
Average Area [0-56]	All sensor readings across all frames *[no sensors activated – all sensors activated for all frames]*
Average Activation [0-65535]	Average number of sensors activated / frame, regardless of pressure *[no activation – all sensors activated to maximum for all frames]*
Peak Activation [0-65535]	Highest value any sensor achieved during a gesture capture *[no activation – all sensors maximally activated for at least one frame]*
Movement Index [0-56]	Quantifies adjacent activity *[no sensors exhibited neighboring activation in previous frame – every sensor's neighbour was activated in the previous frame for all frames]*

Currently, the GRE algorithm calculates features based on available data and evaluates them in various combinations. We use exclusion sets to further tailor the feature behavior to the idiosyncrasies of the platform (de-emphasize known low-probability or commonly misinterpreted gestures and increase likelihood of recognizing common patterns) in a manner similar to a deterministic decision tree. For example, should the average area be large, and there is a high value for average and peak activation, the system may be inclined to recognize a slap – as long as there was a low degree of movement. The use of features refined by exclusion sets has the potential of greater effectiveness than a more purely feature-based approach, but is also somewhat more complex and platform-specific.

While the current GRE is resembles a deterministic decision tree, its next realtime event-based version will introduce probabilistic factors based on previous values of features and gesture probability calculations. A principal advantage of a stochastic (and properly Markovian) GRE is its additional ability to determine gesture output probabilities based on recent past recognitions. For example, if the system has seen several "strokes", a stochastic GRE could on a historical basis be more likely to infer that the current gesture is also a stroke. This characteristic is likely to increase recognition accuracy because it approximates human tendencies in this context, typified by repetitive gestural touch driven by an emotional state that tends to change slowly.

The GRE is most similar to the system described in Knight et al [6], but the contrast is revealed in the systematic differences of the recognized gestures; Knight's gestures are defined primarily by localization – "head-pat", "foot-rub", "side-tickle". In Knight's algorithm, gesture candidates are *first* selected based on location, then subgestures more similar to ours (but not "stroke", the most difficult due to low force activations) are excluded using decision tree logic. In the GRE, the decision tree produces a gesture output from a variety of non-locale-based features. Location is a property of gestures (e.g. a *stroke* that occurred on the back) as opposed to its most defining characteristic (e.g. a side tickle). This leads to different approaches in recognition – more aggressive inferences based on location versus the GRE's reliance on nuanced sensor data. Knight's locale-based gestures and the approach that enables them are relevant in the specific context of an anthropomorphized bear, whereas we sought more general definitions. As well, we infer that Knight's capacitive sensors offer greater resolution and density than the FSRs used here; that is, our algorithm had to operate in a more sparse data environment. Finally, our approach is directed by a near-term need to use Markovian (history-based) processes, whereas other works do not appear to be moving in this direction.

2.2 Current Capabilities

We selected four gestures on the body (*stroke*, *slap*, *poke*, and *pat*) and three on the ears (*pinch*, *pat*, and *squeeze*) as most relevant to the Creature project's goals of studying emotional touch, and required for it to serve as a study platform. Behavioral literature indicates that they are crucial in a context of human-pet or human-child interaction, whereas other important gestures – e.g. *hug* – are not afforded by the robot's lap-situated form. These gestures were not selected for recognition ease.

The GRE is currently tuned to detect these gestures out of a 30-50 frame sequence collected from a fur-covered Creature held on the user's lap. To focus on recognition,

Table 2. Accuracy and error results

a. **Overall accuracy** (w/o uninformative files)

	Pat	Stroke	Slap	Poke	***Avg***
Snout	0%		100%	67%	*56%*
Front	67%	40%	100%	100%	*77%*
Back	25%	17%	0%	50%	*23%*
Rump	0%		0%	50%	*17%*
Side	0%	33%		0%	*11%*

b. **Errors**: % of captures that are *uninformative*

	Pat	Stroke	Slap	Poke	***Avg***
Snout	67%	100%	0%	0%	*42%*
Front	0%	0%	0%	0%	*0%*
Back	20%	0%	80%	67%	*42%*
Rump	0%	100%	20%	0%	*30%*
Side	0%	0%	100%	80%	*45%*

c. **Errors**: % of captures that are *inconclusive*

	Pat	Stroke	Slap	Poke	***Avg***
Snout	0%		0%	0%	*0%*
Front	33%	0%	0%	0%	*8%*
Back	40%	17%	20%	0%	*19%*
Rump	75%		80%	50%	*68%*
Side	100%	0%		0%	*33%*

actuation was turned off for these samples. At present, the GRE uses the entire 2-4 second history for classification, but has the ability to selectively process a shorter time-window (necessary in realtime where multiple concurrent GRE processes must examine varying window lengths to detect gestures of different duration). The current single-process GRE can process a data stream continuously, by defining a fixed-length window that moves along the stream's timeline. The feature values used for the results shown here are listed in Table 1.

Table 2(a) lists recognition success rates for this GRE version from a single user for body-based (non-ear) gestures, while (b-c) show error sources. We indicate where the gesture was applied; classification does not currently indicate area, although this is a minor extension. In Table 2(b-c), where small numbers are good, we define *uninformative data* as samples that exhibited only rest activity, i.e. no activation, generally because the regional sensors were too sparse or insensitive to the lightness of the touch. *Inconclusive data* are samples that exhibited some consistency in non-rest activity that the engine is unable to classify, yet. Gray cells with no value indicate erratic data collected for that region and gesture type, indicating sensor difficulties in snout, rump and side that are exacerbated by the gesture type (stroke, with low force activations; and the very brief slap). Side sensitivity was impacted by loose fur accommodating rib expansion.

The Creature's Front, with the highest and most sensitive coverage due to its low curvature (permitting larger sensors and less bending) produced the highest average recognition rate of 77% (chance=25%), a value which meets our provisional goal of 75% but needs to be verified for usability in interactive contexts.

The remaining areas revealed recognition or sensing difficulties for several region/touch combinations. We interpret *inconclusive* data rates in Table 2(c) as situations with strong potential to be solved with improved features (more, and better optimized), with visually identifiable patterns. With current sensors, the Rump region will be the primary beneficiary of such improvement.

To determine whether *uninformative* data rates of Table 2(b) are a hardware problem, it is useful to consider the pattern of GRE misclassifications. Without space for a full confusion matrix, we summarize. *Pat*⇔*Slap* is the most common, a response to variable sensor sensitivity and positioning. *Slap*⇒*Poke* is a common unidirectional confusion, occurring when a slap hits too-few sensors. *Pat* or *Poke*⇒ *No Detection* is

a too-light pat or a poke that misses a sensor altogether. Finally, *Stroke⇒Pat*, while rare, occurs when only part of a stroke is registered. Together, these patterns confirm a diagnosis of sensor rather than algorithmic weakness.

3 Conclusions and Future Work

In conclusion, we have described a first version of a generalizable engine for recognizing affective touch gestures, embodied in a fully functional, physically animated robot creature, constructed as the fundamental module of a multi-threaded, realtime gesture processor. Our initial assessment shows that the GRE algorithm meets bandwidth and provisional accuracy targets for our experimental purposes when the sensor coverage is adequately dense and sensitive; weaker performance in other areas known to suffer from poorer sensor quality is consistent with the interpretation that sensor network design rather than algorithm poses our current performance bottleneck, and thus is our most immediate target.

In addition to improved sensor system design, our next steps are to move to fully realtime processing by (a) implementing event-driven data stream processing through tying the GRE to an already-existent data event recognizer; and (b) parallelizing multiple moving-window GRE processes to support concurrent consideration of gestures of varying duration. To support more powerful functions in the downstream Emoter module, we will (c) localize gestures, through either logic changes in each of the feature calculator functions or adding a function that weighs relative activation average by per region. Finally, we will fully integrate the GRE with the larger Creature interactive system, as described in [1].

References

1. Yohanan, S., MacLean, K.E.: The Haptic Creature Project: Social Human-Robot Interaction through Affective Touch. In: Proc. of The Reign of Katz and Dogz, 2nd AISB Symp on the Role of Virtual Creatures in a Computerised Society (AISB 2008), Aberdeen, UK, pp. 7–11 (2008)
2. Hertenstein, M.J., Keltner, D., App, B., Bulleit, B., Jaskolka, A.: Touch Communicates Distinct Emotions. Emotion 6, 528–533 (2006)
3. Jones, S.E., Yarbrough, A.E.: A Naturalistic Study of the Meanings of Touch. Communications Monographs 52(1), 19–58 (1985)
4. Dautenhahn, K.: Socially Intelligent Robots: Dimensions of Human–Robot Interaction. Phil. Trans. of the Royal Soc. B: Bio. Sci. 362(1480), 679–704 (2007)
5. Mitsui, T., Shibata, T., Wada, K., Touda, A., Tanie, K.: Psychophysiological Effects by Interaction with Mental Commit Robot. J. of Robotics and Mechatronics 14(1), 13–19 (2002)
6. Knight, H., Toscano, R., Stiehl, W.D., Chang, A., Wang, Y., Breazeal, C.: Real-Time Social Touch Gesture Recognition for Sensate Robots. In: Proc. of IEEE/RSJ Int'l. Conf. on Intelligent Robots and Systems (IROS 2009), St. Louis (2009)
7. Yohanan, S., MacLean, K.E.: A Tool to Study Affective Touch: Goals and Design of the Haptic Creature. In: Proc. of ACM Conf. on Human Factors in Computing Systems (CHI 2009), Works in Progress, pp. 4153–4158 (2009)
8. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall, New Jersey (2009)