Robotic Learning of Haptic Adjectives Through Physical Interaction

Vivian Chu[a,1], Ian McMahon[a,1], Lorenzo Riano[b,1], Craig G. McDonald[a], Qin He[a], Jorge Martinez Perez-Tejada[a], Michael Arrigo[a], Trevor Darrell[b], Katherine J. Kuchenbecker[a]

[a]*Haptics Group, GRASP Laboratory, Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania*
[b]*Department of Electrical Engineering and Computer Sciences, University of California at Berkeley*

Abstract

To perform useful tasks in everyday human environments, robots must be able to both understand and communicate the sensations they experience during haptic interactions with objects. Toward this goal, we augmented the Willow Garage PR2 robot with a pair of SynTouch BioTac sensors to capture rich tactile signals during the execution of four exploratory procedures on 60 household objects. In a parallel experiment, human subjects blindly touched the same objects and selected binary haptic adjectives from a predetermined set of 25 labels. We developed several machine-learning algorithms to discover the meaning of each adjective from the robot's sensory data. The most successful algorithms were those that intelligently combine static and dynamic components of the data recorded during all four exploratory procedures. The best of our approaches produced an average adjective classification F_1 score of 0.77, a score higher than that of an average human subject.

Keywords: robotic perception, haptics, tactile sensing, time-series classification, machine learning

1. Introduction

1.1. Motivation

Manipulation of objects in the real world is a task that goes beyond locating and grasping items of interest. Objects have material properties that need to be properly identified before one can reliably handle them. For example, it is well known that a human attempting to lift an object off a table adjusts his or her grip force and subsequent hand movements based on the coefficient of static friction between their fingertips and the surfaces of the object [1]. Practically speaking, slippery objects need to be grasped more firmly and must be moved less aggressively than sticky objects. When executing a manipulation plan, humans continually predict the tactile signals they will feel and compare their predictions with the actual sensations that occur to monitor their progress and correct any mistakes [2]. To achieve the envisioned benefits of robotic manipulation in human environments [3], robots must develop a similar level of mastery over physical interaction with unknown objects.

Beyond the necessary skill of manipulating everyday objects, robot helpers must also be able to interact smoothly with humans who have little or no technical training. Natural language is likely to be a comfortable communication modality for a wide range of potential users [4, 5, 6]. Like human children, robots will need to be able to learn new words and concepts through direct observation of and interaction with the world. The task of perceptually-grounded language learning requires one to generalize from a small number of examples to deduce the underlying pattern that captures the meaning of the word being learned.

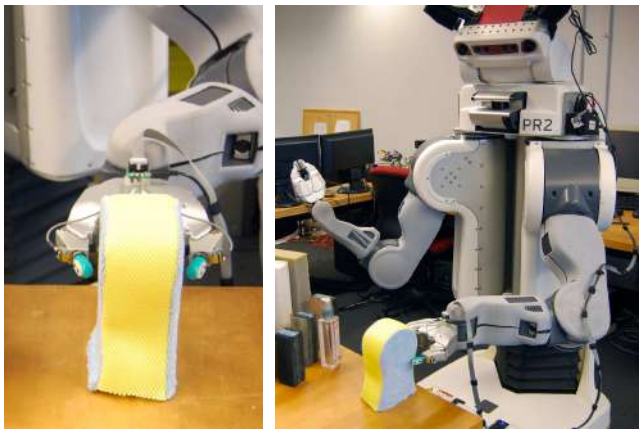

Figure 1: A PR2 robot prepares to touch a car-washing sponge using two BioTac sensors installed in its left hand.

Given the opportunity for robots to function as helpers, we are particularly motivated by the task of learning to describe how objects feel to touch, a challenging undertaking that requires clever physical interaction, rich haptic sensing, and robust machine learning techniques.

One valuable immediate application of a robot that can verbally describe what it feels would be to provide automated and standardized descriptions of physical products such as clothing, stationery, and hand-held electronics. In the age of Internet shopping, online consumers frequently purchase things without knowing what they will feel like to touch. Consumers are less likely to purchase products in this situation, particularly products that have a strong tactile component [7]. One way to address this issue is to provide consumers with scores from in-

[1]These authors contributed equally to this work

dustry experts who have rated these objects using metrics such as KES-FB and FAST [8], which attempt to quantify the tactile properties that people prefer. These metrics are designed for internal product reviews and are not as useful for the average consumer as a detailed and unbiased verbal description of the product's feel. Such a system would allow consumers to search directly for the haptic adjectives they would like a product to have, such as *soft*, *smooth*, and *nice*. Ideally, these labels could be learned automatically from direct physical examples and applied to new products in an impartial manner, a task that is perfect for robotic technology. Haptic adjective descriptions have previously been explored in humans [9], but the use of a robot to perform this task is largely novel.

To investigate the feasibility of robotic learning of haptic adjectives through physical interaction, we created the system pictured in Fig. 1. We sought to enable the depicted robot to autonomously explore objects with its sensorized fingertips and report back an appropriate set of descriptive haptic adjectives. This goal was accomplished by conducting and analyzing two experiments: one in which the robot felt a wide variety of household objects, and another to discover which haptic adjectives humans used to describe these same objects. We developed new methods for processing the heterogeneous and multimodal time-varying information generated by each interaction, and we tested several different techniques for learning the associations between the robot's haptic signals and the human-generated haptic adjective labels.

1.2. Related Work

Humans are capable of haptically recognizing familiar three-dimensional objects through direct contact in just one or two seconds with near 100% accuracy [10]. Individuals accomplish this impressive feat by taking advantage of a rich array of tactile and kinesthetic cues including contact location, pressure, stretch, vibration, temperature, finger position, finger velocity, and muscle force [10, 2]. We similarly believe that tactile cues are essential to enabling a robot to acquire language and execute manual tasks with high accuracy. Furthermore, the sense of touch is inherently interactive – how you move your hand significantly affects the tactile sensations you feel. Interestingly, humans adopt consistent "exploratory procedures" (EPs) when asked to make judgments about specific object properties [11]. For example, the EP of lateral fingertip motion reveals surface texture, pressing shows its hardness, and static contact discloses temperature and thermal conductivity. We believe robots need a corresponding set of exploratory procedures to best interrogate the objects they encounter.

Many robotics researchers have leveraged insights about the human sense of touch to improve robotic manipulation. Early work by Okamura et al. [12] presented robot fingers that roll and slide over the surface of an object to determine texture, ridges, and grooves. Romano et al. [13] imitated human tactile sensing channels and reactions to enable a robot with simple tactile sensors (Willow Garage PR2) to pick up, move around, and set down unknown objects without dropping or crushing them. In a more perception-focused effort, Chitta et al. [14] used the same robot to deduce the identity and contents of a set of beverage containers through touch alone. For the similar task of discriminating containers from non-containers, Griffith et al. [15] demonstrated that the best classification results are achieved when the robot executes a large number of interactions on the target object while attending to diverse sources of sensory data including sight, sound, and touch. Similarly, Sinapov et al. [16] had a custom robot equipped with a vibration-sensitive fingernail use various scratching motions to recognize and categorize everyday textures, with the best results coming from the use of several motions [16]. Oddo et al. used a different novel tactile sensor to achieve good accuracy in discriminating surface roughness [17]. More recently, Fishel and Loeb used a novel biomimetic sensor (SynTouch BioTac) to interact with a library of 117 everyday textures, achieving 95.4% accuracy in identification through the use of cleverly selected tactile features and Bayesian techniques for choosing the most useful movements to perform [18]. Other valuable work on the robotic use of tactile sensors also tends to focus on recognizing particular object instances [19, 20], a task that is related to but distinct from our goal of learning haptic adjectives.

The specific task of quantifying tactile sensations has been explored in the domain of product evaluation for items such as fabric [8] and skin cream [21]. To aid researchers in interpreting the results of opinion studies, there have been a few attempts to develop custom systems to quantify these sensations using machine-learning techniques, e.g, [22, 8]. However, these past approaches have typically used single sensor inputs, which cannot match the richness of human tactile sensation. More recently, Shao et al. showed that accurately evaluating the feel of different packaging materials requires many different channels of touch perception [23]. The need for multi-modal sensing to automatically identify different textures was also noted in the work by Fishel and Loeb [18].

Since the data from haptic sensors usually streams over time, time-series analysis has often been employed to extract information from signals recorded during physical interactions. For example, haptic signals collected from an artificial skin over a robot's entire body were successfully clustered and categorized into human-robot interaction modalities [24]. Similarly, a robot learned to segment tasks by classifying time-series data obtained from an accelerometer and a camera [25]. Our own analysis of time-series data is based on and draws upon the literature in speech recognition, most notably the use of Hidden Markov Models [26].

1.3. Relation to Previous Research

This article builds on preliminary research we published as a short workshop paper [27] and a conference paper [28]. The research reported here significantly extends and refines our prior work, particularly by including more objects, running a formal human-subject study to obtain adjective labels, intelligently merging the traditional approach to feature extraction with our novel HMM-based approach, rigorously and consistently evaluating classifier performance, and thoroughly interpreting all results.

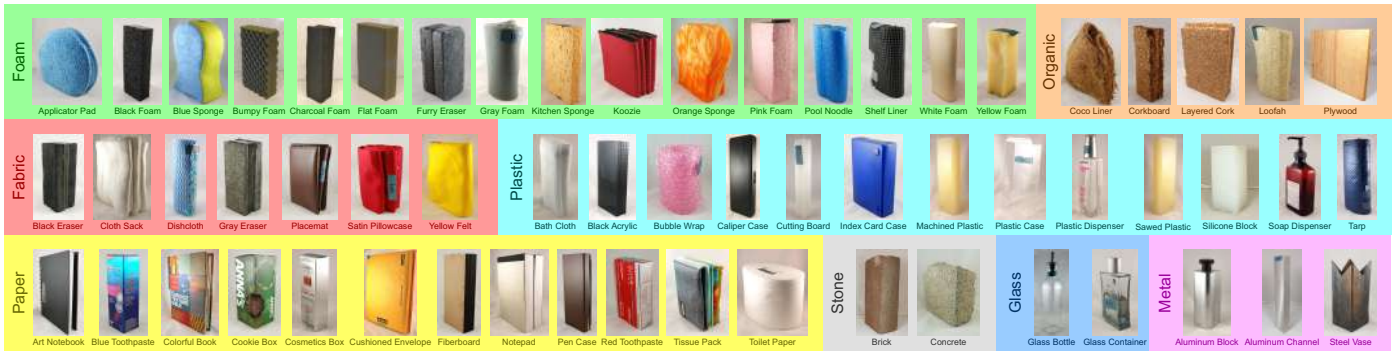

Figure 2: The 60 objects included in the Penn Haptic Adjective Corpus 2, organized by their primary material. Many are household items, and the rest are made from raw materials. Some objects were folded, layered, and/or cut to meet the size and stability requirements for robotic interaction.

2. Materials

2.1. Robotic Hardware

To best emulate what humans experience when exploring objects, we mounted state-of-the-art sensors capable of detecting a wide array of tactile signals onto a humanoid-robotic platform. Specifically, we developed a method to custom-install two Syntouch BioTacs (biomimetic tactile sensors) in the gripper of a Willow Garage PR2 (Personal Robot 2). The modification was successfully performed on PR2s at both Penn and UC Berkeley. Full details on the integration and the software interface packages can be found in our previous work [28] as well as on our project website [29]. The source code for all the approaches described in this paper, including the code to interface with the PR2 and the Syntouch sensors, is available online[2].

The PR2 has two 7-degree-of-freedom (7-DoF) arms that each terminate with a 1-DoF parallel jaw gripper. One BioTac is attached to the distal end of each finger of the left gripper to allow the robot to experience rich tactile sensations as it encloses, squeezes, and strokes objects. The setup can be seen in Fig. 1. Each human-fingertip-sized BioTac is comprised of a rigid core surrounded by conductive fluid within a flexible silicone skin [18]. The BioTac provides measurements of low-frequency fluid pressure (P_{DC}), high-frequency fluid vibrations (P_{AC}), core temperature (T_{DC}), core temperature change (T_{AC}), and a set of nineteen electrode impedances ($E_1 \ldots E_{19}$) spatially distributed across its heated core. Data is read from both fingers at 100 Hz; each packet includes 22 P_{AC} readings and one reading of each other type.

2.2. Objects

The enhanced PR2 interacted with 60 objects that represent a wide range of physical properties within the limits of what the robot can grasp. As seen in Fig. 2, each object is able to stand upright on a table and has two flat, parallel, vertical sides with identical surface properties. To fit between the BioTacs, each object is between 1.5 cm and 8 cm thick. To allow the PR2 to slide vertically down the object without colliding with the table, the objects have a minimum height of 10 cm. No objects are sharp, pointed, cold, or hot to avoid damaging the BioTacs. Furthermore, all included objects are clean, dry, and durable to prevent damage to the system's electronics and to maintain a consistent feel over time.

3. Experimental Setup

To understand and generalize the experience of touching objects, we need knowledge of both how an object feels and how to describe those sensations. A touch-sensitive robot can show us precisely through bits and bytes what it has sensed, but it has no means to describe what impression the experience has made. On the other hand, humans can describe their perception of an object in words, but they cannot precisely share what they felt.

We developed two parallel experiments to capture the information necessary to learn the meaning of a representative set of haptic adjectives. Both the augmented PR2 robot and a group of human subjects touched the 60 objects shown in Fig. 2 under carefully controlled conditions that included no visual feedback, no auditory feedback, and no prior knowledge of the objects. As described below, the two experiments were designed to give the robot and the human subjects as similar an experience as possible.

3.1. Robotic Exploration

For the PR2 to interact with the objects, we programmed a custom controller in ROS (Robot Operating System) to perform a fixed set of exploratory movements. As described in [28], the controller is fully integrated with ROS tabletop object detection and can use motion-planning algorithms to autonomously place the robot gripper around the object. However, to speed up data collection, we programmed the robot to move its open gripper to a fixed point above the table. The target object was then placed on the table within the gripper with a small random perturbation in pose designed to mimic the uncertainty of autonomous object perception. This process is necessary only for rapid data acquisition, and the tabletop object detection pipeline can be used for the general case when the object position and shape are unknown. Moreover, as introduced in [28], the controller has an initial probing phase where it adjusts the gripper

[2]https://github.com/imcmahon01/Penn-haptics-bolt

position so that the object is centered between the fingers. Because all of the tested objects have parallel sides, the robot's two fingers contact the object equally.

Once the gripper is around the object, the PR2 begins the following sequence of four Exploratory Procedures (EPs): *Squeeze*, *Hold*, *Slow Slide*, and *Fast Slide*. We designed these EPs to match a subset of the archetypal movements humans make when touching objects to discern their haptic properties [11]. The human EPs of Pressure, Enclosure, Static Contact, and Lateral Motion were selected from Lederman and Klatzky's full list [11] because they can be reliably executed by our two-fingered robot. For the robotic EPs based on Lateral Motion, we also drew inspiration from Fishel and Loeb's work on using a BioTac for texture recognition [18]. A full description of each EP can be found in [28].

Before the robot begins executing the EPs, it performs a *Center* action to equalize the contact pressure between the two BioTacs. For this step, the PR2 repeatedly closes its gripper until it feels contact, opens, and moves its gripper a small amount toward the finger that first touched the object. This state ends when both fingers contact the object at approximately the same time. The PR2 then performs a high velocity *Tap* to confirm the centered location of the object. Although we previously identified it as an EP, here we omit data from this very brief *Tap* interaction because it yields little unique information about the object and was not based on a human exploratory procedure. The gripper opens by a small amount when *Tap* completes.

The first EP, *Squeeze*, was designed to match the human EP of Pressure, which is used to discern object hardness [11]. Here, the gripper closes at a constant slow velocity to a specific P_{DC} value before opening at the same constant velocity until no contact is felt on either finger. The controller then transitions to *Hold*, which was designed to correspond to the human EPs of Enclosure (global shape and volume) and Static Contact (temperature) [11]. After closing to a proportion of the full depth reached during the previous state, the robot gently holds the object for ten seconds to let the heated fingers reach thermal equilibrium with the object, which begins at room temperature. The final two EPs have the robot move its hand downwards while lightly holding the object, creating sliding interactions that match the human EP of Lateral Motion [11]. Similar to [18], *Slow Slide* uses a stronger contact with a downward speed of 1 cm/s, while *Fast Slide* has half the contact strength and a speed of 2.5 cm/s.

This whole procedure was performed ten times on each of the 60 objects, with slight repositionings each time, resulting in 600 robotic data collection trials. Each robotic trial took approximately 90 seconds to complete. As done for our preliminary dataset of 510 trials [28], these results were saved as ROS "bagfiles" that contain time histories of many PR2 sensors. The work reported in this paper uses only the PR2 gripper position, the PR2 gripper aperture, all readings from both BioTac sensors, and the timing of the controller's states and sub-states.

3.2. Human Subject Study

To understand how people describe their haptic interactions with the same set of 60 objects, we conducted an experiment with thirty-six human subjects. All procedures were approved by the University of Pennsylvania's Institutional Review Board under protocol #816464. Subjects gave informed consent and were compensated $15 for participation.

Figure 3: The 25 adjectives selected for both the humans and the robot to use in describing each object. Each adjective's text is stylized here to convey its meaning, but they were all presented to the human subjects in a standard font. "Gritty" is marked with an asterisk because the human subjects did not agree on a single positive example of this adjective, so it is omitted from further discussion in this paper.

The primary objective of this study was to obtain human-verified labels of a range of haptic adjectives for each of the objects in our corpus. Our preliminary work on this project [28] tested a list of 34 adjectives in a pilot study that included only four participants. Analysis of the responses given by these subjects showed that certain adjectives in this list were not being used at all, while others were highly correlated with one another. The irrelevant and redundant adjectives were removed to yield the list of 25 haptic adjectives shown in Figure 3.

The study participants consisted of 34 right-handed and 2 left-handed individuals, 10 male and 26 female, ranging in age from 18 to 21 years. All were college students with normal upper-limb motion and hand function. During the experiment, the subject was seated at a table in front of a vertical panel. The subject could reach around the panel but could not see their hand or the object. Each object was held in the air by a ring stand to prevent the participant from touching the table, which might have served as a tactile reference. This restriction also beneficially averted the human tendency to lift and move the object, which was prohibited because the augmented PR2 is not capable of lifting all of the selected objects. A three-part survey was presented on a computer display visible to the subject. All participant responses were provided verbally and recorded by the experimenters and a video camera. During the experiment, the subject wore noise-cancellation headphones playing white noise to hide sounds produced by interactions with the objects as well as ambient noises. These headphones were also used to communicate verbally with the subject throughout experiment.

Subjects were instructed to use only their index finger and thumb to interact with the presented objects. Four exploratory procedures were allowed: pressure, enclosure, static contact, and lateral movement. In order to mimic the robot's interactions with the objects, subjects were not allowed to scratch, lift, twist, or perform any other action on the object. They were also told to avoid contact with the object's edges and adjacent surfaces.

Every participant was presented twice with 20 objects, which were evenly chosen from the 60 objects in the corpus. Each object was thus touched by approximately twelve individu-

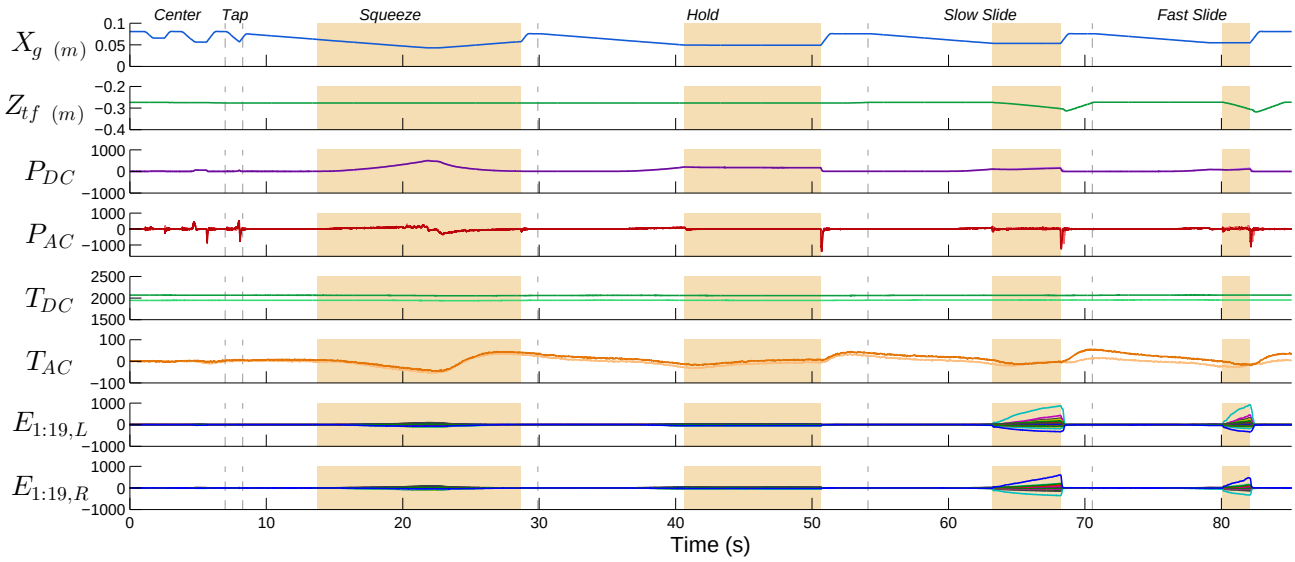

Figure 4: These haptic signals were collected from one trial of the PR2 robot exploring the Blue Sponge object shown in Fig. 1. The recording is automatically segmented into the exploratory procedure phases using the robot controller states. All BioTac sensor reading are in numerical units of the twelve-bit ADC. These BioTac signals are centered at zero during non-contact, with the exception of the temperature signal, T_{DC}.

als. During the first randomly ordered presentation, the subject commented freely on the feel of each object. This step served to familiarize the individual with the range of objects and the testing procedure. We had the subject touch a compliant stress ball between trials to attempt to wash out their tactile memory of the previous object. The second randomly ordered presentation of the objects involved both binary and Likert ratings of pre-determined haptic adjectives. A randomly ordered list of the 25 haptic adjectives shown in Fig. 3 was first shown on the screen, and the participant selected the words they considered characteristic of the object. There was no minimum or maximum number of adjectives to be selected. Immediately following this verbal checklist, the subject rated the same object on ten basic haptic adjectives using a five-point Likert scale. This final step was included to test hypotheses about specific antonym relationships among common haptic adjectives and is thus not considered in this paper.

3.3. Penn Haptic Adjective Corpus 2

Together, the results of both experiments constitute the Penn Haptic Adjective Corpus 2 (PHAC-2). This corpus contains all data from the 600 trials collected during the robot experiment as well as all of the labels collected during the human subject study. PHAC-2 is a significant improvement over the original Penn Haptic Adjective Corpus 1 (PHAC-1) because it includes nine more objects and three times as many human-provided labels per object. Furthermore, the PHAC-2 adjective labels are likely to be higher quality than those in PHAC-1 because they were collected with more standardized procedures involving subjects who were completely naive to the research.

4. Experimental Data

4.1. Robot

We extracted several relevant haptic signals from the bagfiles recorded during the robot's interactions with the 60 PHAC-2 objects. Figure 4 displays the PR2 and BioTac data from a single interaction with the object Blue Sponge. Coming from the robot, the signal X_g shows the gripper aperture (distance between the fingertips), where a reading of 0.1 m means the gripper is completely open. Z_{tf} is the vertical position of the gripper with respect to the robot's torso coordinate frame. The additional five sets of signals were gathered from the BioTac sensors themselves. Each of the pressure (P_{DC}, P_{AC}) and temperature (T_{DC}, T_{AC}) graphs contains both left finger (light signal) and right finger (dark signal) readings. The nineteen spatially distributed electrodes in each finger are presented separately to aid interpretation.

Before the robot's fingers have come into contact with any object, we calculate and store the mean of every BioTac signal except core temperature, T_{DC}. These means are subtracted from subsequent measurements in that run to mitigate sensor drift and discrepancies between the two BioTacs. These zeroed signals are made available to the machine learning pipeline in addition to the unmodified versions of the signals. During a trial, each EP's signal is automatically segmented to include only the portion of the movement containing physical interaction with the object (highlighted in Figure 4). The duration of the reduced signal varies by movement and by object. These interesting controller subsections are passed to the machine learning pipeline for feature extraction.

4.2. Human

Each of the 36 individuals who participated in the human subject study provided 25 binary adjective ratings for each of

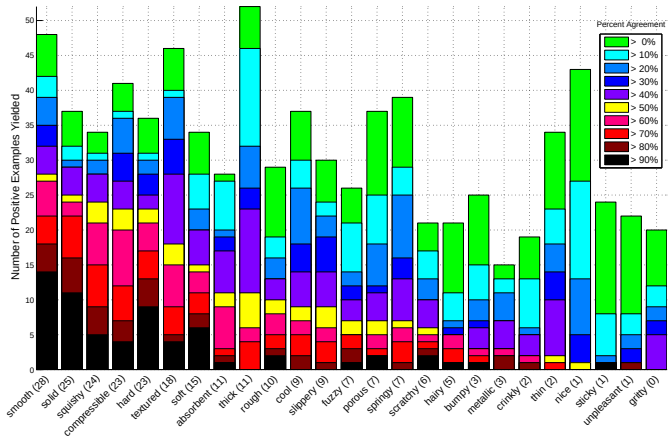

Figure 5: The number of objects that subjects identified as positive examples for each of the 25 adjectives. The colored levels indicate the proportion of the subjects choosing a given object for a given adjective. For example, only one of the sixty objects (Kitchen Sponge) was labeled *absorbent* by more than 90% of the subjects who touched it, but 28 total objects were labeled *absorbent* by at least one subject. We selected a threshold of 50% to provide the ground-truth haptic adjective labels for the sixty objects. The adjectives are arranged in decreasing order of positive examples, as determined by the 50% majority.

the 20 objects they touched, yielding a total of about 12 human interactions per object. In this data set, a rating of 1 signifies that the subject thought the object exhibited the qualities of the given adjective. For instance, subject 17 said the Blue Sponge was *springy*, *soft*, *squishy*, and *thick*, rating all of the other adjectives a binary 0 for this object. The number of adjectives selected for a given object varied from subject to subject and also from object to object within the same subject; the minimum was 0, the maximum was 11, and the median was 5.

Fig. 5 shows the number of objects that subjects selected as being characteristic of each adjective, with a breakdown of the percent agreement. Given these binary vote counts, we had to devise a uniform method of choosing an overall binary label for a given object-adjective pair. The human votes were tallied and then normalized by the total number of individuals who felt that object. If more than 50% of the human subjects labeled an object-adjective pair positive, the overall majority vote was labeled positive. Under this paradigm, the Blue Sponge shown in Fig. 1 was labeled *absorbent*, *compressible*, *soft*, and *squishy*. Interestingly, the adjective *gritty* was not applied to any object by more than 50% of the human subjects, so it was omitted from consideration for robotic adjective learning. Furthermore, three other adjectives (*nice*, *sticky*, and *unpleasant*) were found to apply to only one object each (Applicator Pad, Silicone Block, and Coco Liner, respectively), leaving no opportunity for the robot to learn to generalize the meaning of these words within this set of objects.

4.3. Train and Test Sets

All of the selected robot interaction data and all of the majority-vote haptic adjective labels were combined into a database for use by our machine learning algorithms. To allow us to not only train our algorithms but also test their performance on new interaction data, we devised a method for dividing the collected database into train and test sets. We aim to create a separate binary classifier for each of the adjectives listed in Figure 3, so we needed twenty-four different train/test splits (omitting *gritty*, which did not apply to any object). Each of these train/test sets contains all 60 objects, since any given object was labeled as either a positive or negative example of each adjective by the majority vote.

Each adjective's split between training and testing examples was based on the positive and negative majority-vote labels. These splits were constructed by gathering at least 10% of all positively labeled and at least 10% of all negatively labeled objects into the testing set for each adjective. The remaining ~90% of positive and negative objects were placed into the training set. Notice that all 10 runs of a given object were kept on the same side of the train/test split to avoid fitting classifiers to an object rather than the underlying adjective. The number of objects in each testing set can be determined by evaluating the following equation twice, once for all positive examples and once for all negative examples:

$$N_{\text{test}} = \lceil 0.1 \cdot \gamma \rceil \tag{1}$$

where γ is the total number of positive or negative examples for this particular adjective. After the size of each set was determined, the members of each set were randomly selected without replacement from the available options. This procedure guarantees the presence of at least one positive and at least one negative example in the testing set for every adjective. To give a concrete example of our method for choosing train/test splits, consider the adjective *absorbent*, which has 11 positive object examples and 49 negative object examples. Equation (1) yields a test set containing 2 positive *absorbent* object examples and 5 negative examples. The remaining 9 positive objects and 44 negative objects form the training set.

We set up a special case to handle the three adjectives that have only one positive object example: *nice*, *sticky* and *unpleasant*. Since it was impossible to create whole positive training and testing sets, we split these objects by trial. Five randomly selected trials for the adjective's positive object were designated for the training set and five for testing.

5. Features

While the human adjective labels are simple to comprehend, the robot recorded a voluminous quantity of diverse data during each of its 600 trials, as exhibited in Fig. 4. Thus, we had to find good ways to select a small number of numerical values (features) to represent each channel and EP of the interaction for use in our machine learning algorithms. We approached feature selection from the two complementary viewpoints of (1) calculating static values using hand-crafted formulas and (2) computing novel dynamic values using automatic signal processing.

Investigation of both types of features is well motivated by the fact that humans attend to and combine a variety of sensory cues when judging the haptic properties of objects [11]. For example, Tan et al. conducted a clever set of experiments to determine whether subjects touching virtual objects through rigid

Table 1: P_{DC} Feature Equations

Maximum	Mean	Greatest Slope Change
$\max(P_{DC})$	$\frac{1}{n}\sum_{i=1}^{n}(P_{DC})$	$\max([P'_{DC}(n) - P'_{DC}(n-1)])$

plates base their estimates of compliance on the actual compliance of the surface, the terminal force felt during the interaction, and/or the mechanical work done by their fingers [30]. The precise mechanisms of human haptic perception are not known for all possible object properties, nor do we assume that the augmented PR2's sensory capabilities are perfectly matched to those of a human, so we sought to create a wide variety of potentially useful features. Our static method follows previous research on robotic texture categorization such as Fishel and Loeb [18] and Sinapov et al. [16], where features are carefully designed based on knowledge of human and robotic touch to capture salient characteristics of the signal. Our dynamic method takes a very different approach and aims to model the way the entire haptic signal changes over time.

5.1. Static Feature Calculation

Our static method calculates single-valued features from the robot's tactile and kinesthetic sensations; as discussed in Section 4.1, all BioTac signals except T_{DC} are zeroed at the start of each trial. Features are then extracted from the most relevant time segments of the four EPs: *Squeeze*, *Hold*, *Slow Slide*, and *Fast Slide*. The selected segments are highlighted in Figure 4. Each of the features described below is calculated for each BioTac for each of the four EPs. This segmentation by EP allows us to analyze the impact of each EP on learning a particular adjective and also ensures that each EP has equal representation in the input.

Low-Frequency Fluid Pressure (P_{DC}). The first two features are the maximum and mean of P_{DC} during the EP segment, chosen to capture the finger's internal pressure change. The third feature finds the greatest change in a smoothed version of the P_{DC} signal's slope over time. The equations for these three features are shown in Table 1. These static features were selected because of the noticeable difference in P_{DC} signal size and shape between stiff and compliant objects.

High-Frequency Fluid Vibrations (P_{AC}). Recent work by Fishel and Loeb showed that frequency-domain analyses of BioTac contact vibrations are very successful for texture classification tasks [18]. To obtain our features, we first converted the 2200 Hz P_{AC} temporal vibration signal into a non-normalized energy spectral density, *ESD*, where ω is the vector of frequencies, as follows:

$$ESD(\omega) = \frac{1}{2\pi}\left|\int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,\mathrm{d}t\right|^2 = \frac{1}{2\pi}F(\omega)F^*(\omega) \qquad (2)$$

To represent the *ESD* via single-valued features, we computed the total energy of the *ESD* curve, plus the spectral centroid (SC), variance (SV), skewness (SS), and kurtosis (SK) using the equations shown in Table 2. A specific run of the two sliding

Table 2: P_{AC} Feature Equations

Feature	Equation
E_{total}	$\frac{\omega_{\max}-\omega_{\min}}{2N}\sum_{k=1}^{N}[ESD(\omega_k) + ESD(\omega_{k+1})]$
SC	$\frac{\sum ESD(\omega)\cdot\omega}{\sum ESD(\omega)}$
SV	$\frac{\sum ESD(\omega)\cdot(\omega - SC)^2}{\sum ESD(\omega)}$
SS	$\frac{\sum ESD(\omega)\cdot(\omega - SC)^3}{SV^{\frac{3}{2}}\cdot\sum ESD(\omega)}$
SK	$\frac{\sum ESD(\omega)\cdot(\omega - SV)^4}{SV^2\cdot\sum ESD(\omega)} - 3$

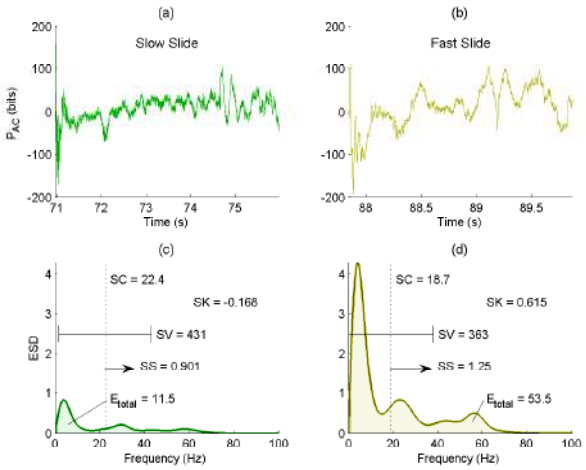

Figure 6: Features created from the raw P_{AC} signals from both *Slow Slide* and *Fast Slide* from an exploration of the object Car Sponge. (a) and (b) show the raw signals from a single BioTac. (c) and (d) show the conversion to the frequency domain and the values pulled out for each of the five features.

EPs can be seen converted from raw P_{AC} to the selected static features in Figure 6.

Core Temperature (T_{DC}) & Core Temperature Change (T_{AC}). The BioTacs are internally heated to above human body temperature, so they transfer heat to any room-temperature surfaces they come into contact with. We referred to work by Lin et al. [31] for guidance on processing the temperature channels of the BioTacs; their investigations showed that T_{DC} and T_{AC} can capture the thermal conductance of different materials. The rate at which the heat is transferred varies between objects of different material and geometry. We capture this transfer rate by calculating the total area under the T_{AC} curve and the time constant τ of an exponential fit of T_{DC} over time. The equations for these static features can be found in Table 3.

Electrode Impedances ($E_1 ... E_{19}$). The BioTac's 19 electrodes give spatial information about the contact between the finger and the objects. Because the electrodes are all connected to the same volume of conductive fluid inside the finger, we found that many of the electrode signals are highly correlated with one another. We used this knowledge to lower the dimensionality of the 19 electrodes to the two most significant principal

Table 3: Thermal Feature Equations

Area Under T_{AC}	T_{DC} Curve Profile
$\frac{t_{\max}-t_{\min}}{2N}\sum_{k=1}^{N-1}[T_{AC}(t_k)+T_{AC}(t_{k+1})]$	$p_1 + p_2\, e^{-\frac{t}{\tau}}$

Table 4: EP-based PCA Variance for the 19 Electrodes

EP	First Component	Second Component
Squeeze	91.7%	3.4%
Hold	91.7%	3.2%
Slow Slide	79.7%	11.3%
Fast Slide	84.7%	6.9%

Table 5: Robot Feature Equations

Max. Aperture	Min. Aperture	Height Range
$\max(X_g)$	$\min(X_g)$	$\max(Z_{tf}) - \min(Z_{tf})$

components using Principal Component Analysis (PCA) [32]. PCA was performed separately on the 19 channels of data from each EP to capture the unique characteristics of each type of interaction. The top two components were found to capture the majority of the variance, as can be seen in Table 4. Once the 19 electrodes have been reduced to their two highest principal components, we fit a fifth-order polynomial to each component's coefficient over time. The 12 coefficients of the two polynomial equations are our static features for the electrodes.

Robot Channels. The first two features that depend directly on the robot's movement are the maximum and minimum of the PR2's gripper aperture (X_g) during the selected EP, values that change dramatically with object thickness and compliance. The third static feature based on robot motion is the range of the gripper's vertical position (Z_{tf}). This feature was selected to capture the amount of friction between the BioTacs and the object, since the robot hand tracks its desired position more closely when the object is slippery. The corresponding equations are shown in Table 5.

These methods give us 47 static features per EP per trial, broken down as 22 for each BioTac and three from the PR2.

5.2. Dynamic Feature Calculation

Our dynamic features seek to encapsulate how the robot's tactile sensing signals change over time. Here we consider the P_{DC}, P_{AC}, T_{AC}, and $E_1 \ldots E_{19}$ BioTac readings over all four EPs, omitting T_{DC}, X_g, and Z_{tf} because they largely correlate with the chosen tactile readings. As done in our prior work [28], we used Hidden Markov Models (HMMs) [26] to capture the variations in the data. Since HMMs rely on an alphabet of finite symbols, the first step was to create an adaptive discretization pipeline that converts the raw haptic signals into symbols. The following three-step pre-processing pipeline takes as inputs the signals coming from both fingers as an $n \times d$ matrix, where n is the number of time steps in the recording and d is the number of channels for a given sensor times two fingers.

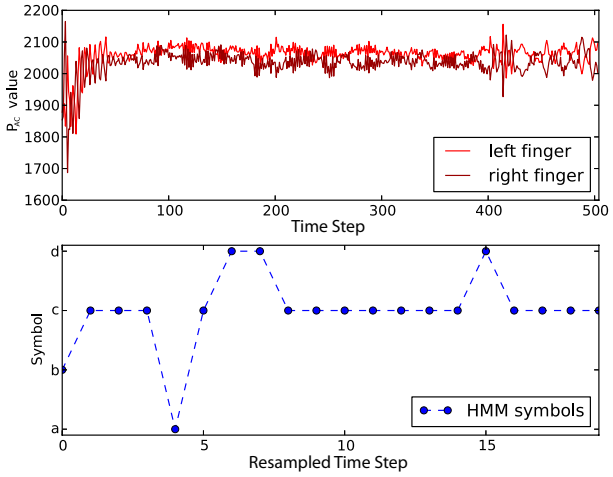

Figure 7: An example of the vector quantization produced by the HMM pipeline. The upper graph shows the signals of two P_{AC} sensors on the right and left fingers. The lower graph represents the corresponding symbols produced by vector quantization. This example used only 4 of the available 5 symbols in the alphabet.

1. *PCA*: Principal Component Analysis [32] reduces the dimensionality of the input data while retaining a percentage of the variance. We calculated the minimum number of new dimensions r such that 97% of the input's variance was retained. Our experiments revealed that even for the 36-dimensional vector corresponding to the two fingers' electrodes, as few as 4 dimensions were sufficient to represent the data. The output of PCA is a $n \times r$ matrix.
2. *Resample*: An HMM calculates the probability of observing a sequence of symbols. This product of probabilities becomes lower as the length of the input sequence increases. To avoid numerical instability and over-fitting, we resampled the PCA data to an $m \ll n$ dimensional vector using linear interpolation. The output is an $m \times r$ matrix.
3. *Discretization*: Once an input signal is reduced in dimensions, we use vector quantization [33] to convert it to discrete symbols drawn from an s-sized alphabet. In other words, we treated the $m \times r$ input matrix as m independent r-dimensional vectors and performed K-Means to identify s clusters. The output is a m-dimensional vector of discrete elements.

Figure 7 shows the effects of the above pipeline on a sample pair of P_{AC} signals recorded when the robot performed *Slow Slide* on the Art Notebook. The output is a 20-dimensional vector of discrete symbols drawn from an alphabet of cardinality 5. Only 4 of the 5 available symbols were used by this particular time series. Whenever a significant change happens in the input data, there is a corresponding change of symbol. The pipeline is otherwise very robust to noise.

Training the HMMs. We train our HMMs using the Baum-Welch algorithm with the discretized signals as input. As summarized in Table 6, the discretization pipeline and the HMMs have hyper-parameters that strongly influence their behavior. To identify the best values for these parameters, we split the

Table 6: Parameters to train the HMMs.

Parameter	Values	Description
h	$[12, 15, 18]$	Number of hidden components with the HMMs
m	$[20, 25, 30]$	Resampling size
s	$[5, 10, 12]$	Number of clusters in K-Means

training data set into 2/3 for initial training and the remaining 1/3 for training evaluation. We used cross-validation to select the parameters that yield the greatest log-probability over a sequence. Each HMM is trained to recognize a time series coming from the positive examples of a single haptic adjective for a given exploratory procedure and a given tactile sensor channel. Consequently, we trained 16 HMMs (four motions times four sensors) for each adjective. During feature extraction, each trial is associated with a 16-dimensional vector of log-probabilities, interpreted as a vector of dynamic features.

5.3. Feature and Label Analysis

Before passing them into any machine learning algorithm, we analyzed our static and dynamic features by looking at where our 60 objects lie in each feature space. This analysis allows us to determine whether the selected features correspond well to physical properties of the objects.

The static feature vector has 188 elements (47×4), and its values do not depend on the adjective being considered. Given the way the HMMs are trained, there is a distinct 16-element dynamic feature vector for each adjective, so we chose a representative adjective (*hard*) to report here. The features were first centered, normalized, and averaged across each of the 10 trials per object, yielding 60 object-specific average feature vectors for both the static and dynamic approaches. These average feature vectors were passed into PCA, and the top two principal components are shown in Fig. 8 (static) and Fig. 9 (dynamic). In both cases, neighboring objects seem to share physical properties such as compliance and texture, giving us confidence that both approaches are producing valuable features. Note also that the two spaces resemble one another but are not identical; the static and dynamic features capture somewhat different aspects of the interactions.

In parallel, we used correspondence analysis (CA) to investigate the validity of the human adjective ratings. CA allows one to understand the relationship between stimuli (for us, objects) and descriptors (for us, binary haptic adjectives from majority voting) within a dataset that is categorical, non-negative, and uniformly scaled [34]. CA places both the objects and the adjectives in a shared perceptual space where the distance between elements represents their perceived dissimilarity [34]. Figure 10 shows where each adjective and each object sit in the first two dimensions yielded by CA. Examination of the plot gives credibility to the human-provided adjective labels. For example, we see that objects that are *metallic* and *cool* sit close together, and objects such as Loofah and Gray Eraser are near *rough* and *scratchy*.

We can now compare the PCA plots of the static and dynamic data features with the CA plot of the adjective labels. Immediately we see that all three top principal components seem to correspond to compliance, a highly salient object property that is fundamental to the adjective *hard*. However, when we start to look at the second component, it is more difficult to make a clear correlation between the three feature spaces. While it is apparent that the adjective space's second dimension relates to object roughness, the second component of the two feature spaces only slightly correlate with roughness, leading us to speculate

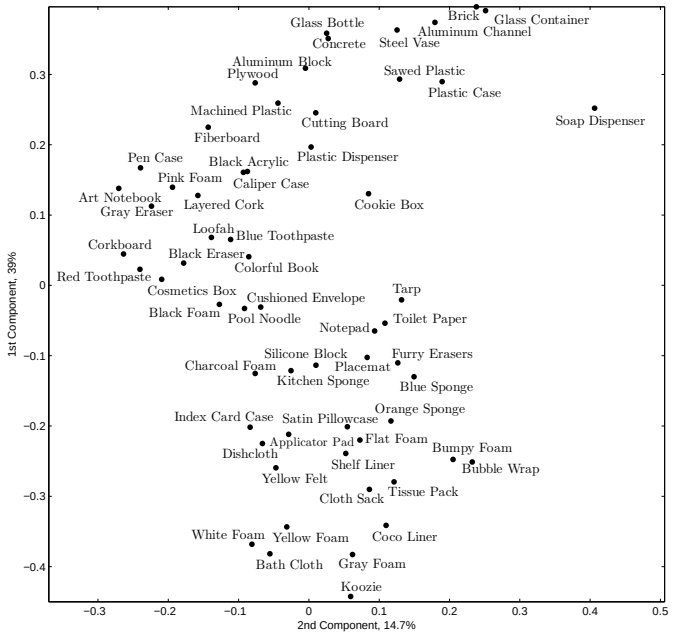

Figure 8: The first two principal components of the static feature space; each PHAC-2 object is shown at the average location occupied by its trials.

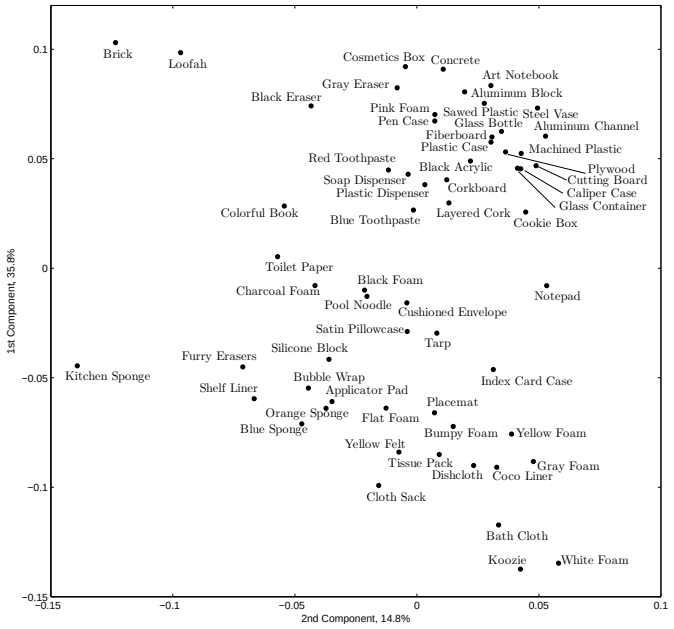

Figure 9: The first two principal components of the dynamic feature space for the representative adjective *hard*; each PHAC-2 object is shown at the average location occupied by its trials.

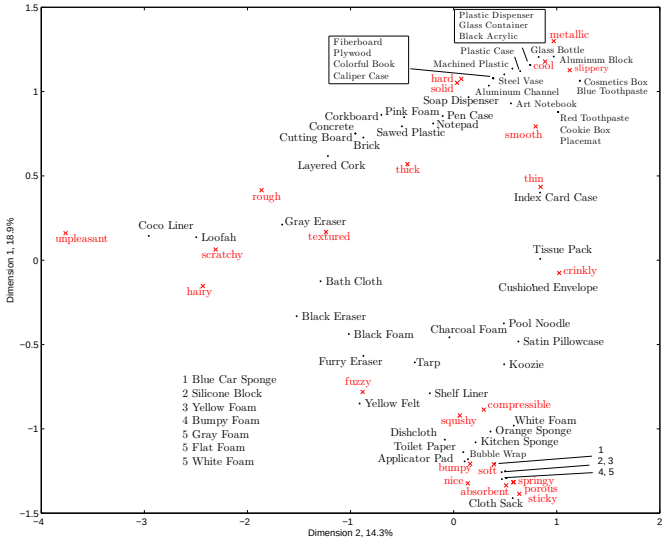

Figure 10: The first two dimensions of the adjective-object space created from correspondence analysis (CA)

that our robot may not be able to feel this object property as easily as humans can.

6. Training the Classifiers

We trained classifiers to capture the static and dynamic nature of the data using the database of 90% training object splits for each adjective, holding back the 10% of objects designated for testing each adjective, as described in Section 4.3.

6.1. Separate Static and Dynamic Classifiers

Given the static and dynamic features introduced in Section 5, we trained two corresponding groups of 24 adjective-specific classifiers using each adjective's designated training set (both positive and negative examples). Each classifier used a linear support-vector machine (SVM) with the L2 metric [35]. We reserved 3 objects (30 feature vectors) from each training set for use in validation. The F1 score of each classifier was calculated by averaging over 100 randomly selected validation sets. We used cross-validation and grid search to select the SVM error penalty factor C, with values ranging from 10^0 to 10^6. This approach yielded 24 adjective-specific classifiers based on the 188-dimensional static features (Sec. 5.1) and 24 based on the 16-dimensional dynamic features (Sec. 5.2).

6.2. Static + Dynamic Classifiers

We explored two techniques for merging the static and dynamic feature information. For the first method, we simply combined the static and dynamic features to create a 204-dimensional feature vector. We then trained a linear SVM using the approach described in the section above. We call this classifier *Combined*. The second approach merged the two kernels rather than the feature vectors. Given a valid kernel for the static features K_s and a valid kernel for the dynamic features K_d, it is well known that a linear combination of the two is still

Table 7: Metric Equations

Precision	Recall	F_1
$\frac{tp}{tp+fp}$	$\frac{tp}{tp+fn}$	$2 \cdot \frac{precision \cdot recall}{precision + recall}$

a reproducing kernel for the underlying Hilbert space. We can therefore construct a new merged kernel K_m as:

$$K_m = \alpha K_s + (1 - \alpha) K_d \quad (3)$$

where K_s and K_d are both linear kernels, and $0 \leq \alpha \leq 1$. After conceiving of this approach, we trained an SVM for each adjective using cross-validation and grid search to find the penalty factor C and the mixing coefficient α. We call this multi-kernel learning classifier *MKL* (see [36] for similar approaches.)

7. Results

All of the results reported below were obtained through testing on the reserved adjective-specific test sets, which were never seen during training, as described in Section 4.3. The training and testing were performed on Linux-based PC computers with Intel i7 processors with a single core speed of 3.8 GHz and total RAM of 16.0 GB. On these machines, a single object exploration can be classified with all adjectives in approximately 10 seconds. The metrics that we have selected are precision, recall, and F_1 score, as defined in Table 7, where tp is the true positive count (the number of correct positive results returned by the classifier), fp is false positives (the number of results incorrectly labeled positive by the classifier), and fn is false negatives (the number of positive results missed by the classifier).

Table 8 provides a summary of the results averaged over all adjectives for each of the four classifiers introduced in Section 6, plus two additional EP-specific classifiers discussed below. Importantly, we have removed the contributions from *nice*, *sticky*, and *unpleasant* whenever calculating the mean over all adjectives because these labels have just one positive example in the entire corpus. As explained in Sec. 4.3, the classifiers for these three adjectives were trained on half of the trials recorded for their exemplar object and then tested on the other half; naturally, the machine learning algorithms overfit to these specific objects and yielded test scores that do not accurately represent how well the classifiers generalize.

7.1. Exploratory Procedure Breakdown

We first looked at how classifiers trained on the features from only an individual EP (*Squeeze*, *Hold*, *Slow Slide*, *Fast Slide*) performed across all of the adjectives. Each was trained using the same methods described in Sec. 6.1 but looking at only the 47 static features or only the 4 dynamic features constructed from the designated EP's data. The results are shown in Table 9 and Table 10, with adjectives listed in decreasing order of the number of positive examples in the training set. When looking at an individual phase, the static method generally performs better than the dynamic method, presumably because it has more

Table 8: Average scores for the developed adjective classifiers.

Classifier	Precision	Recall	F_1
Static, Individual EP	0.40	0.29	0.31
Dynamic, Individual EP	0.20	0.08	0.11
Static	0.48	0.51	0.47
Dynamic	0.62	0.47	0.52
Combined	0.61	0.59	0.57
MKL	0.79	0.76	0.77

Note that each value reported here is the mean of the individual scores, so the mathematical relationship between precision, recall, and F_1 does not necessarily follow the equation given in Table 7.

Table 9: F_1 Score Across Adjectives and EPs Using Only Static Features

	Squeeze	*Hold*	*Slow Slide*	*Fast Slide*	PE*
smooth	0.658	0.508	0.421	0.122	25
solid	0.846	0.893	0.968	0.833	22
squishy	0.774	0.877	0.655	0.712	21
compressible	0.769	0.909	0.868	0.776	20
hard	0.667	0.630	0.967	0.952	20
textured	0.000	0.095	0.087	0.065	16
soft	0.744	0.378	0.500	0.564	13
absorbent	0.667	0.529	0.850	0.600	9
rough	0.000	0.182	0.706	0.571	9
thick	0.250	0.000	0.160	0.000	9
cool	0.250	0.583	0.133	0.286	8
slippery	0.167	0.000	0.000	0.222	8
fuzzy	0.000	0.000	0.000	0.000	6
porous	0.696	0.526	0.258	0.000	6
springy	0.235	0.000	0.250	0.091	6
scratchy	0.000	0.000	0.000	0.000	5
hairy	0.000	0.333	0.250	0.533	4
bumpy	0.000	0.000	0.143	0.182	2
metallic	0.133	0.667	0.000	0.000	2
crinkly	0.000	0.000	0.000	0.000	1
thin	0.000	0.000	0.000	0.000	1
nice	0.000	0.000	0.909	0.571	0.5
sticky	0.000	0.000	0.000	0.000	0.5
unpleasant	0.444	0.000	0.615	1.000	0.5

*PE indicates the number of positive examples in the training set.

information available. Overall, using only a single EP fails to do well at identifying adjectives when the number of positive examples falls below about 10. There are some interesting exceptions with the static method for adjectives such as *hairy* and *metallic* that have EPs that do relatively well even with only a small number of positive examples, seemingly because the EP is particularly well suited to judging that adjective's characteristics. On the other end of the spectrum, there also exist adjectives that do poorly even with a high number of positive examples, notably *textured* and *smooth*.

7.2. Static and Dynamic Classifiers

We next looked at the performance of the full static and dynamic classifiers, which consider data from all four EPs; their average scores are reported in the third and fourth rows of Table 8. Interestingly, these results show that the full dynamic classifier achieves a somewhat higher overall score than the full static classifier, reversing the trend found for individual EPs. Figure 11 shows the differences between the scores for each of the 24 adjectives. The dynamic classifier completely fails to generalize for *springy*, *rough*, *thick*, *textured*, *fuzzy* and *slippery*, but these zero scores are largely offset by high scores for the other adjectives. In contrast, the static classifier fails to generalize only for *scratchy*, *crinkly* and *thin*, but the scores for the remaining adjectives remain low, thus explaining its overall lower performance when compared to the dynamic classifier.

Table 10: F_1 Score Across Adjectives and EPs Using Only Dynamic Features

	Squeeze	*Hold*	*Slow Slide*	*Fast Slide*	PE*
smooth	0.738	0.432	0.185	0.125	25
solid	0.836	0.815	0.681	0.750	22
squishy	0.776	0.636	0.278	0.571	21
compressible	0.750	0.776	0.065	0.235	20
hard	0.846	0.720	0.235	0.571	20
textured	0.000	0.000	0.000	0.000	16
soft	0.667	0.182	0.083	0.000	13
absorbent	0.788	0.000	0.000	0.000	9
rough	0.000	0.000	0.000	0.000	9
thick	0.000	0.000	0.000	0.000	9
cool	0.000	0.000	0.000	0.000	8
slippery	0.000	0.000	0.000	0.000	8
fuzzy	0.000	0.000	0.000	0.000	6
porous	0.000	0.000	0.000	0.000	6
springy	0.000	0.000	0.000	0.000	6
scratchy	0.000	0.000	0.000	0.000	5
hairy	0.000	0.000	0.000	0.000	4
bumpy	0.000	0.000	0.000	0.000	2
metallic	0.000	0.000	0.000	0.000	2
crinkly	0.000	0.000	0.000	0.000	1
thin	0.947	0.000	0.000	0.000	1
nice	0.000	1.000	0.000	0.000	0.5
sticky	0.000	0.000	0.000	0.333	0.5
unpleasant	1.000	0.000	1.000	0.000	0.5

*PE indicates the number of positive examples in the training set.

7.3. Merged Classifiers

The results for the static and dynamic classifiers show that the two are complementary in their capabilities. We therefore expect a classifier that merges the static and dynamic information to perform better than either individual classifier. When we conducted these experiments, we discovered that only the MKL classifier showed a significant improvement over the individual approaches, as seen in the bottom two rows of Table 8. Figure 11 shows a side-by-side comparison of the adjective-specific performance for the Combined and MKL classifiers. The MKL classifier dominates the results, with an average F_1 score of 0.77. Surprisingly, both classifiers failed to learn the adjective *scratchy*, which was at least partially identifiable by the dynamic classifier.

7.4. Humans

To obtain a baseline by which to judge the robot's performance, we re-analyzed the results of the human-subject study by treating each individual as a classifier. Each participant's binary predictions were scored against the human majority label and then aggregated across objects using the same precision, recall, and F_1 metrics used to quantify the robot's performance. Figure 12 presents a box plot of this F_1 score per adjective for all human subjects. The MKL classifier is overlaid on top of this figure for a side-by-side comparison of robot and human performance. In general, the robot's MKL classifier tends to score closer to the majority labels than the average human, achieving an average F_1 across all adjectives of 0.77 versus the human average F_1 score of 0.65.

8. Discussion

The presented results show that adjective classifiers trained on more heterogeneous data outperform classifiers trained on

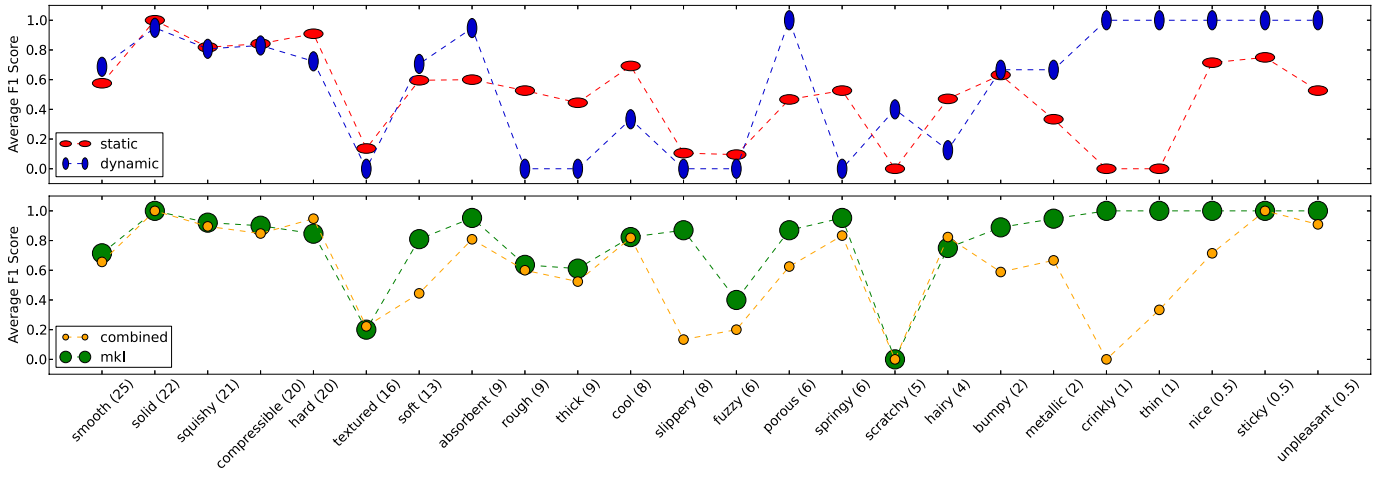

Figure 11: Comparisons of the F_1 scores between the Static and Dynamic classifiers (top row) and Combined and MKL classifiers (bottom row). The parenthetical numbers with the adjective labels are the number of positive examples in the training set.

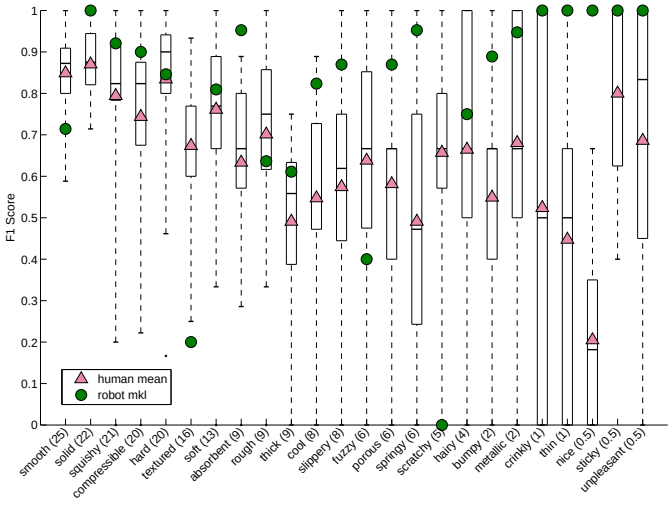

Figure 12: F_1 score comparison between the robot's MKL classifier and all 36 human subjects (when tested against the majority-voted adjective labels). Each box plot shows the second and third quartiles of the human F_1 scores, the line shows the median, the filled symbol shows the mean, the whiskers show the range up to 1.5 times the inter-quartile range, and dots mark outliers.

only a subset of the data when attempting to label objects that have never before been touched. Although this result is not new in the machine-learning community, it has profound implications in haptics and robotics in general.

The first source of heterogeneity comes from the different ways in which the robot physically explored the object. While certain EP-specific static classifiers do reasonably well with a small number of positive examples, the F_1 scores are very low overall (Table 9), and the EP-specific dynamic classifiers do even worse. This trend shows that when a robot is learning the rich characteristics of an object, a single motion often does not yield enough useful information to be reliable at generalizing adjectives. The first jump in classification scores occurs when the information from all the motions is unified into a single feature vector (either static or dynamic), explaining the difference between the low individual EP scores and the higher full static or dynamic scores seen in Table 8.

The BioTac sensors offer a multi-modal view into the haptic perceptual space, with information coming from motion, pressure, temperature, shape, and vibration. Our experiments show that our two feature extraction approaches behave differently when trying to recognize different adjectives. Figure 11 shows that some dynamic classifiers fail at recognizing adjectives that the static classifiers otherwise can, and vice versa. The best choice is therefore unifying the two approaches to extracting information from the voluminous sensory data streams.

The Combined classifier's results indeed prove this point, albeit with an overall F_1 score (0.57) that barely outperforms that of the Dynamic classifier (0.52). A close look at Figure 11 shows that, although most of the 0 scores have been raised, *crinkly* and *scratchy* still have a 0 score even though their Dynamic classifier score was 1. Also adjectives like *textured*, *fuzzy* and *slippery* had low scores in both the dynamic and static classifiers, and this trend has been kept in the combined classifier. In contrast, the MKL classifier seems to have learned how to use the available data best, achieving an average F_1 score of 0.77, which is better than the average human participant. A close examination of the right part of the lower plot in Figure 11 shows that MKL learned to generalize better than the Combined classifier when presented with a small number of positive examples.

The behavior of the MKL classifier can be further analyzed by comparing the mixing factor α with the scores of the Static and Dynamic classifiers, as shown in Figure 13. According to equation (3), the greater α is, the greater is the weight assigned to the static features. Adjectives like *hairy*, *solid*, *absorbent* and *cool* have non-extreme α values, indicating that both static and dynamic features played an equal role in characterizing the adjective. On the other side, adjectives like *porous* and *thin* are chosen using only either the static or the dynamic kernels, choices that are consistent with the relative scores of the two individual classifiers. The adjectives *slippery* and *hard* represent

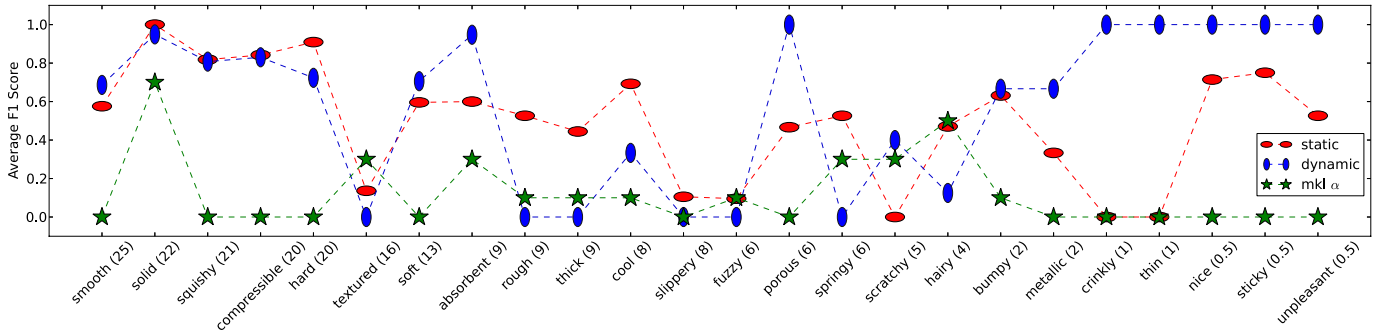

Figure 13: Mixing factor α for the MKL classifier in relation to the scores of the Static and Dynamic classifiers. An α value of 1 indicates only the Static kernel is considered, while a value of 0 indicates that only the Dynamic kernel is considered.

an exception to this trend, where a greater weight is given to the kernel whose corresponding classifier performs worse.

The MKL approach is adept at combining the best of the static and dynamic feature sets, yielding better overall results than either classifier alone. An interesting parallel exists between the MKL classifiers and humans: it is known that humans can judge object compliance using only tactile cues or only kinesthetic cues, with tactile cues being more useful [37]. Like the robot combining static and dynamic features, human performance at compliance discrimination is best when both tactile and kinesthetic cues are available [37].

Overall, the responses of the average human matched the majority vote with a similar level of performance as the MKL classifier, validating that MKL is indeed learning the mapping between physical interaction signals and haptic adjectives. In fact, the MKL classifier is somewhat better than the average human at predicting whether a new object will be voted as a positive adjective example, a result that substantiates the value of multi-modal tactile sensing and heterogenous exploratory procedures.

In examining the data for the single positive examples of *nice*, *sticky*, and *unpleasant*, we noticed that the PR2 was more consistent than expected across trials. All of our developed classifiers could adeptly predict the labels for the withheld trials, leading us to remove these scores from our aggregate results to avoid bias. This finding suggests that future versions of this experiment probably do not require 10 robotic interaction trials for each object; approximately five trials would be sufficient.

Considering all of the results, it was reassuring that adjectives trained with more positive examples generally achieved a higher F_1 score during testing. A notable exception was with the adjective *textured*, which had 16 positive examples but achieved an F_1 score of only 0.2. This discrepancy led us to look more closely at all of the other texture-related adjectives, which include *smooth*, *rough*, *fuzzy*, *scratchy*, and possibly *absorbent*. It seems that these adjectives generally did not perform as well as other adjectives that had a similar number of training examples. These results support the findings from Section 5.3 that texture was not the first or second principal component of our feature space; it seems that our robot cannot feel textures as easily as humans can, most likely due to wear of the patterned ridges on the BioTac skins.

In this work we chose linear models for both feature extraction (linear PCA) and classification (linear kernels). The main reason behind this choice is the substantial computational advantage linear models have over more complicated kernels. This decision in turn drastically reduces the training time when searching in a large parameter space via cross-validation. The results in Table 4 also support our choice, as most of the variance in the sensors is captured by the first component of the PCA projection of the data.

9. Conclusion and Future Work

We set out to create a robotic system capable of touching everyday objects and describing them with haptic adjectives. We performed an experiment to learn what words humans choose to describe a large set of selected objects, and we collected haptic data from a robot that touched these same objects ten times each. The richness of the signals collected from the BioTac sensors enabled us to perform a multi-modal analysis of object properties. Based on prior robotics research and knowledge of human touch, our primary hypothesis was that characterizing the feel of everyday objects with the acuity of human perception requires information gathered from different kinds of interactions and diverse sensors. We therefore constructed both traditional static and novel dynamic features from the haptic data for use in learning the meaning of the haptic adjective labels. We built classifiers of increasing sophistication and tested them on previously unfelt objects. The results we have obtained are very encouraging, in that a multi-kernel classifier trained on both static and dynamic features performed better than the average human subject on the adjective labeling task. These results could not be obtained by looking at one source of information alone. Our experiments therefore prove our initial hypothesis and pave the road for other new approaches to tactile information analysis and classification.

There are many avenues available for future work. This research calculated a large number of static and dynamic features from the robot data; it would be useful to analyze which features were most important for the learning of individual adjectives and for overall performance. It would also be interesting to apply a Bayesian selection approach similar to that done by

Fishel and Loeb [18] to choose the best exploratory procedures for discriminating each adjective. This approach should work well because our static classifier seemed to do best when the EP matched the characteristics of the adjective; as seen in Table 9, both *cool* and *metallic* had the best results with *Hold*, the EP that provides the most information about the object's thermal properties, while both *hairy* and *rough* favored sliding EPs, which we would expect to capture the most information about texture.

This research did not focus on dexterous object interaction but rather used a parallel-jaw gripper to perform EP's. To provide similar tactile experiences, we required the human subjects to emulate the capabilities of the robot's two-fingered gripper. It would be interesting to reverse this paradigm and have the humans interact with objects freely, lifting and sliding the objects as they desired. One could then investigate a multi-fingered BioTac-enhanced robotic hand capable of performing movements similar to those of unrestricted humans. This approach would require a more complex manipulation controller for lifting and placing objects, but it would allow one to study a broader range of adjectives and evaluate the effectiveness of the other EP's described by Lederman and Klatzky [11].

10. Acknowledgements

The authors thank J. Nappo for helping find many of the objects in Figure 2. We thank N. Fitter for guiding us in the use of Correspondence Analysis. This work was supported by the Defense Advanced Research Projects Agency (DARPA) in the United States as part of Activity E within the Broad Operational Language Translation (BOLT) program, as well as by the University of Pennsylvania and the University of California, Berkeley.

11. Vitae

Vivian Chu received the B.S. degree in Electrical Engineering and Computer Science from the University of California, Berkeley, in 2009 and the M.S.E. degree in Robotics from the University of Pennsylvania in 2013. She will start her Ph.D. degree in Robotics at the Georgia Institute of Technology in the fall of 2013. From 2009 and 2011, Vivian worked in the database research team at IBM Research, Almaden, where she focused on natural language processing (NLP) and intelligent information integration. At Penn, she worked in Katherine Kuchenbecker's Haptics Research Group in the GRASP Lab, where she focused on grounding language through object manipulation and tactile sensing. Her research interests include multi-modal sensor integration, NLP, and applying machine learning techniques for robotic learning in unstructured environments.

Ian McMahon received his B.S. degree in Computer Engineering from the Pennsylvania State University in 2009 and his M.S.E. degree in Robotics from the University of Pennsylvania in 2012. During his masters degree, he worked in Katherine Kuchenbecker's Haptics Research Group in the GRASP Lab. His research interests include haptic sensing and computer vision for use in mobile manipulation, as well as human-robot interaction. Ian is currently working at the startup Parametric Dining LLC as a Machine Learning Engineer.

Lorenzo Riano received the first class honors master degree in Computer Science at the University of Napoli "Federico II", Italy, and his Ph.D. in robotics at the University of Palermo, Italy. He is a Research Scientist at the University of California, Berkeley. Previously he was a Research Associate at the Intelligent Systems Research Centre, University of Ulster, UK. His main research interests are in robotics, machine learning and computer vision.

Craig G. McDonald received the B.S.E. degree in Mechanical Engineering and Applied Mechanics from the University of Pennsylvania in 2012. He will start his Ph.D. degree in Mechanical Engineering at Rice University in the fall of 2013. At Penn, he worked in Katherine Kuchenbecker's Haptics Research Group in the GRASP Lab. At Rice, he is working in Marcia O'Malley's MAHI Lab, as well as NASA Johnson Space Center's Dexterous Robotics Lab, where he will focus on the design and control of robotic exoskeletons. His research interests also include haptic interfaces, robotic manipulation, and machine learning.

Qin He received her B.S. degree in Mechanical Engineering from Shanghai Jiao Tong University in 2012, and began her masters degree in Robotics at the University of Pennsylvania in 2012. During her master's degree, she worked in Katherine Kuchenbecker's Haptics Research Group in the GRASP Lab. Currently she is a research assistant in Daniel Lee's group in the GRASP Lab, working on the DARPA Robotics Challenge. Her research interests include haptic sensing and learning algorithms in robotics.

Jorge Martinez Perez-Tejada is a B.S.E. candidate with majors in Electrical Engineering and Cognitive Science at the University of Pennsylvania, where he is expected to graduate in May 2014. Jorge volunteered in the GRASP Lab starting in 2011, and he joined Katherine Kuchenbecker's Haptics Research Group in the summer of 2012. In spring 2013, he did independent research on haptic perception under Prof. Kuchenbecker's supervision. In the fall of 2013, he will join the Computational Memory Lab at Penn to work on his senior thesis.

Michael Arrigo is a master's student in the Linguistics department at the University of Pennsylvania, and his area of focus is computational linguistics. He received his B.A. degree in Cognitive Science and Linguistics from the University of Pennsylvania in 2013. In 2012, he joined Katherine Kuchenbecker's Haptics Research Group in the GRASP Lab, where he gathered data on human haptic adjectives as a model for the PR2 robot behavior. His research interests include natural language processing and applying linguistic theory to language and speech technology.

Trevor Darrell is head of the Computer Vision Group at the International Computer Science Institute and is on the faculty of the CS Division at UC Berkeley. His group develops algorithms to enable multimodal conversation with robots and mobile devices, and methods for object and activity recognition on such platforms. His interests include computer vision, machine learning, computer graphics, and perception-based human computer interfaces. Prof. Darrell was on the faculty of the MIT EECS department from 1999 to 2008, where he directed the Vision Interface Group. He was a member of the research staff at Interval Research Corporation from 1996-1999, and he received the S.M. and Ph.D. degrees from MIT in 1992 and 1996, respectively. He obtained the B.S.E. degree from the University of Pennsylvania in 1988, having started his career in computer vision as an undergraduate researcher in Ruzena Bajcsy's GRASP lab.

Katherine J. Kuchenbecker earned the B.S., M.S., and Ph.D. degrees in Mechanical Engineering from Stanford University in 2000, 2002, and 2006, respectively. She worked as a postdoctoral research associate at the Johns Hopkins University from 2006 to 2007. At present, she is an Associate Professor in Mechanical Engineering and Applied Mechanics at the University of Pennsylvania. Her research centers on the design and control of haptic interfaces, and she directs the Penn Haptics Group, which is part of the GRASP Laboratory. Prof. Kuchenbecker serves on the program committee for the IEEE Haptics Symposium and other conferences in the fields of haptics and robotics, and she has won several awards for her research, including an NSF CAREER award in 2009, inclusion in the Popular Science Brilliant 10 in 2010, and the IEEE Robotics and Automation Society Academic Early Career Award in 2012.

References

[1] G. Cadoret, A. M. Smith, Friction, not texture, dictates grip forces used during object manipulation, Journal of Neurophysiology 75 (1996) 1963–1969.

[2] R. S. Johansson, J. R. Flanagan, Coding and use of tactile signals from the fingertips in object manipulation tasks, Nature Reviews Neuroscience 10 (2009) 345–359.

[3] C. C. Kemp, A. Edsinger, E. Torres-Jara, Challenges for robot manipulation in human environments: Developing robots that perform useful work in everyday settings, IEEE Robotics and Automation Magazine (2007) 20–29.

[4] D. J. Brooks, C. Lignos, C. Finucane, M. S. Medvedev, I. Perera, V. Raman, H. Kress-Gazit, M. Marcus, H. A. Yanco, Make It So: Continuous, Flexible Natural Language Interaction with an Autonomous Robot, AAAI Technical Report WS-12-07 (2012) 2–8.

[5] J. F. Gorostiza, M. A. Salichs, End-user programming of a social robot by dialog, Robotics and Autonomous Systems 59 (2011) 1102–1114.

[6] X. Chen, J. Ji, J. Jiang, G. Jin, F. Wang, J. Xie, Developing high-level cognitive functions for service robots, in: Proc. 2010 International Conference on Autonomous Agents and Multiagent Systems, pp. 989–996.

[7] A. V. Citrin, D. E. S. Jr., E. R. Spangenberg, M. J. Clark, Consumer need for tactile input: An Internet retailing challenge, Journal of Business Research 56 (2003) 915–922.

[8] N. Pan, Quantification and evaluation of human tactile sense towards fabrics, International Journal of Design and Nature 1 (2006) 48–60.

[9] M. Hollins, S. Bensmaïa, K. Karlof, F. Young, Individual differences in perceptual space for tactile textures: Evidence from multidimensional scaling, Attention, Perception, & Psychophysics 62 (2000) 1534–1544.

[10] R. L. Klatzky, S. J. Lederman, V. A. Metzger, Identifying objects by touch: An "expert system", Perception and Psychophysics 37 (1985) 299–302.

[11] S. J. Lederman, R. L. Klatzky, Extracting object properties through haptic exploration, Acta Psychologica 84 (1993) 29–40.

[12] A. M. Okamura, M. L. Turner, M. R. Cutkosky, Haptic exploration of objects with rolling and sliding, in: Proc. 1997 IEEE International Conference on Robotics and Automation, volume 3, pp. 2485–2490.

[13] J. M. Romano, K. Hsiao, G. Niemeyer, S. Chitta, K. J. Kuchenbecker, Human-inspired robotic grasp control with tactile sensing, IEEE Transactions on Robotics 27 (2011) 1067–1079.

[14] S. Chitta, J. Sturm, M. Piccoli, W. Burgard, Tactile sensing for mobile manipulation, IEEE Transactions on Robotics (2011) 558–568.

[15] S. Griffith, J. Sinapov, V. Sukhoy, A. Stoytchev, A behavior-grounded approach to forming object categories: Separating containers from noncontainers, IEEE Transactions on Autonomous Mental Development (2012) 54–69.

[16] J. Sinapov, V. Sukhoy, R. Sahai, A. Stoytchev, Vibrotactile recognition and categorization of surfaces by a humanoid robot, IEEE Transactions on Robotics 27 (2011) 488–497.

[17] C. Oddo, M. Controzzi, L. Beccai, C. Cipriani, M. Carrozza, Roughness encoding for discrimination of surfaces in artificial active-touch, IEEE Transactions on Robotics 27 (2011) 522–533.

[18] J. A. Fishel, G. E. Loeb, Bayesian exploration for intelligent identification of textures, Frontiers in Neurorobotics 6 (2012) 1–20.

[19] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, W. Burgard, Object identification with tactile sensors using bag-of-features, in: Proc. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 243–248.

[20] N. Gorges, S. E. Navarro, D. Goger, H. Wörn, Haptic object recognition using passive joints and haptic key features, in: Proc. 2010 IEEE International Conference on Robotics and Automation, pp. 2349–2355.

[21] M. Loden, H. Olsson, L. Skare, T. Axell, A. H. Ab, Instrumental and sensory evaluation of the frictional response of the skin following a single application of five moisturizing creams, Journal of the Society of Cosmetic Chemists 43 (1992) 13–20.

[22] Z. Du, W. Yu, Y. Yu, L. Chen, Y. Ni, J. Luo, X. Yu, Fabric handle clusters based on fuzzy clustering algorithm, in: Proc. 2008 IEEE International Conference on Fuzzy Systems and Knowledge Discovery, volume 3, pp. 546–550.

[23] F. Shao, X. Chen, C. Barnes, B. Henson, A novel tactile sensation measurement system for qualifying touch perception, Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine 224 (2010) 97–105.

[24] T. Taichi, M. Takahiro, I. Hiroshi, H. Norihiro, Automatic categorization of haptic interactions-what are the typical haptic interactions between a human and a robot?, in: Proc. 2006 IEEE-RAS International Conference on Humanoid Robots, pp. 490–496.

[25] S. Lenser, M. Veloso, Non-parametric time series classification, in: Proc. 2005 IEEE International Conference on Robotics and Automation, pp. 3918–3923.

[26] L. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition, Proceedings of the IEEE 77 (1989) 257–286.

[27] I. McMahon, V. Chu, L. Riano, C. G. McDonald, Q. He, J. M. Perez-Tejada, M. Arrigo, N. Fitter, J. Nappo, T. Darrell, K. J. Kuchenbecker, Robotic learning of haptic adjectives through physical interaction, in: Proc. 2012 IEEE/RSJ IROS Workshop on Advances in Tactile Sensing and Touch-based Human-robot Interaction, Vilamoura, Algarve, Portugal.

[28] V. Chu, I. McMahon, L. Riano, C. G. McDonald, Q. He, J. M. Perez-Tejada, M. Arrigo, N. Fitter, J. C. Nappo, T. Darrell, K. J. Kuchenbecker, Using robotic exploratory procedures to learn the meaning of haptic adjectives, in: Proc. 2013 IEEE Internatational Conference of Robotics and Automation, p. TBA.

[29] Penn Haptics, http://bolt-haptics.seas.upenn.edu/, 2013.

[30] H. Z. Tan, N. I. Durlach, G. L. Beauregard, M. A. Srinivasan, Manual discrimination of compliance using active pinch grasp: The roles of force and work cues, Perception and Psychophysics 57 (1995) 495–510.

[31] C. H. Lin, T. Erickson, J. Fishel, N. Wettels, G. Loeb, Signal processing and fabrication of a biomimetic tactile sensor array with thermal, force and microvibration modalities, in: Proc. 2009 IEEE International Conference on Robotics and Biomimetics, pp. 129 –134.

[32] I. Jolliffe, Principal component analysis, Wiley Online Library, 2005.

[33] S. Lloyd, Least squares quantization in PCM, IEEE Transactions on Information Theory 28 (1982) 129–137.

[34] D. Picard, C. Dacremont, D. Valentin, A. Giboreau, Perceptual dimensions of tactile textures, Acta Psychologica 114 (2003) 165–184.

[35] C. Burges, A tutorial on support vector machines for pattern recognition, Data mining and knowledge discovery 2 (1998) 121–167.

[36] M. Gönen, E. Alpaydın, Multiple kernel learning algorithms, Journal of Machine Learning Research 12 (2011) 2211–2268.

[37] W. M. B. Tiest, A. M. Kappers, Kinaesthetic and cutaneous contributions to the perception of compressibility, in: M. Ferre (Ed.), Proc. 2008 EuroHaptics, volume 5024 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin Heidelberg, 2008, pp. 255–264.