Learning to Identify Container Contents Through Tactile Vibration Signatures

Carolyn L. Chen[1], Jeffrey O. Snyder[2], and Peter J. Ramadge[3]

***Abstract*— We examine using a simple contact sensor coupled with standard machine learning algorithms to classify and count objects shaken in a container. The contact sensor measures the resulting vibrations, and these signatures are used to learn a classifier that maps vibration signatures to known object categories. A linear support vector machine trained on labeled vibration signatures achieves a mean binary classification accuracy of 99% over 66 pairs of objects and a mean multi-class classification accuracy of 94% over 12 classes. It is also shown that useful tasks such as approximate counting of objects over the range 1 to 10 is possible. We see potential applications of these ideas in service robots engaged in cleanup and inventory control in labs, workshops, stores, warehouses and homes.**

I. Introduction

Tactile sensing is increasingly important in robotic perception, with applications ranging from manipulation to health care. For example, tactile sensing enabled the Willow Garage PR2 robot to grasp unknown delicate objects [1], and haptic probes in autonomous robotic palpation have demonstrated effective localization of subcutaneous inclusions [2], [3]. We posit that to maximize information extraction and subsequent effectiveness, tactile sensors will be closely integrated with machine learning methods. In part, this trend has already begun. For example, [4] and [5] used the Support Vector Machine (SVM) and Naive Bayes classifiers, respectively, to distinguish surfaces using accelerometer vibration measurements acquired as a stylus is dragged across the surface. Similarly, [6] used a multi-sensor commercial tactile sensing device to classify surfaces using a process of Bayesian exploration yielding comparable results.

Motivated by the recent successful investigations of vibrations in [4] and [5], we focus on exploring how well a single contact vibration sensor (accelerometer or contact microphone) can detect differences in the contents of a container (such as a cup or beaker) by shaking it. We benchmark performance using a high-bandwidth studio quality microphone. When the container is shaken, the objects within it collide with the container walls and excite the vibration modes of the container and shaking mechanism. These vibrations can be measured directly on the exterior container surface using an accelerometer or contact microphone, or indirectly by sensing the aggregate sound generated. We call a measurement of these vibrations a *vibration signature* of the container and its contents. Our goal is to determine to what extent a vibration signature characterizes the container contents by using standard techniques for dimensionality reduction, classification, and clustering. The resulting accuracy quantifies how well vibration signatures characterize the container contents.

Our study focuses on the identification of container contents through induced container vibrations. Our contribution goes well beyond demonstrating that this is feasible. Using standard methods such as SVMs, we demonstrate an average pairwise classification accuracy of 99% across 66 object pairs and an average multi-class accuracy of 94% across twelve object classes. This is comparable to results in object classification using more complicated tactile sensor systems [7], [8]. We also show that for some objects it is possible to count the number of items in the container within a tolerance of ± 2 with an accuracy of 99% and to distinguish objects despite variability in the number of items. Furthermore, we reveal the limits of using a simple contact sensor by comparing its performance on challenging tasks with that of a high-bandwidth studio microphone.

II. Prior Work

Vibration analysis has a long history. For example, it has been used to track machine health [9], to monitor structural integrity of beams [10] and railways [11], to analyze pipes [12] and power systems [13], and to determine the characteristics of musical instruments [14]. It has been used to classify the terrain on which an exploratory vehicle is currently traveling [15], [16] or a surface texture that a tool tip is traversing [4], [5]. Indeed, any form of exploratory interaction with an object that induces vibrations (e.g., shaking, poking, or scratching) renders vibration analysis relevant. Despite this, vibration analysis has not been extensively explored for robotic haptic sensing and classification.

Shaking is of particular interest because it is a typical exploratory action of infants and is known to help humans determine object shapes [17]. In [17], a microphone recorded the sound produced by shaking paper materials, plastic bottles containing water, and rigid-bodies with the intent of distinguishing these classes of objects. In another recent study [18], multi-modal sensing (vision, sound and joint-torque in a 7-Degrees of Freedom (DOF) arm) was paired with ten exploratory procedures, shaking being one, to distinguish between glass marbles, rice, beans, and screws in identical plastic containers. The results indicate that shaking combined with analysis of the sound produced was more effective for distinguishing the objects inside the container.

[1] Dept. Electrical Engineering & Computer Science, University of California, Berkeley, CA 94720, USA. carolyn.chen@berkeley.edu

[2] Dept. of Music, Princeton University, Princeton, NJ 08544, USA. josnyder@princeton.edu

[3] Dept. of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA. ramadge@princeton.edu

A few similarities between our study and [17], [18] can be drawn. All of these studies involve shaking-induced vibration analysis and object classification. Both [17] and [18] experiment with clearly distinct objects and use microphones to sense aggregate vibration data. The study in [17] did not attempt to classify objects within containers, and [18] used one plastic container and four easily distinguished objects. In contrast, we focus on a direct contact sensor, examine the impact of the container on classification accuracy, and explore the resolution limits of this exploratory procedure by using similar objects that are difficult to distinguish.

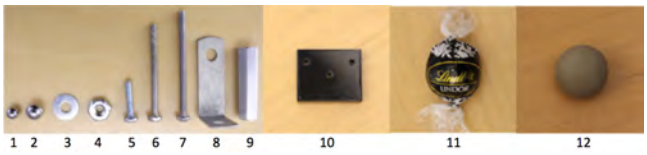
III. Experimental Methodology

We collected data in a consistent manner using a custom shaking mechanism. The slider-crank mechanism was used to translate rotational motion from a DC brushed metal gearmotor to reciprocating motion of a simple clamp end effector. The length of the reciprocating motion and the motor rotation speed were set to 4" and just over 2 Hz, respectively, mimicking how a human might shake a container.

We used three vibration sensors. The two of primary interest are contact sensors: a three-axis accelerometer (LIS344ALH, bandwidth 1.8 kHz [19]) and a contact microphone (measured bandwidth 5 kHz). These were mounted on the container exterior, clamped inside the end effector. For reference purposes, we also used a studio quality Earthworks SR71 cardioid microphone (bandwidth 50 Hz to $>$ 20 kHz) mounted two feet above the shaker. For brevity, we will refer to the studio and contact microphones as the air and contact mic., respectively. The bandwidth of the contact mic. is a potential advantage over the accelerometer. Likewise, the bandwidth of the studio mic. makes it the reference standard on challenging tasks. The signals were digitized at 44.1 kHz and 24 bit floating point precision using a U46 XL interface.

Four containers were tested (glass, plastic, cardboard, paper) along with twelve object types (see Fig. 1). For each object, a set amount (number or weight) was shaken in each container. The container was not removed from the end effector until all objects in that experiment were explored. In each case, we recorded 65 sec of vibration data. We regard the first 5 sec as a transient and the last 60 sec are split into non overlapping T sec vibration signatures for the object-container combination. Using $T \leq 1$ gives $\lfloor 60/T \rfloor$ labelled vibration signatures for each shaking run.

Each T sec vibration signature is a vector $w \in \mathbb{R}^N$ with $N = \lfloor 44100 \times T \rfloor$. We normalize each w to unit 2-norm. Similarly, each accelerometer sample is matrix $A = [A_x, A_y, A_z] \in \mathbb{R}^{N\times 3}$ normalized to have unit 2-norm columns. We then compute the magnitude of the discrete Fourier transform of each signature and keep only the positive frequency indices (frequency length $M = N/2 + 1$ for N even). Finally, the DFT321 method [20] is applied to the accelerometer data. Let $j \in [0, M-1]$ be the frequency index and $A^{(j)}$ denote the j-th row of the normalized A. Then DFT321 takes the 2-norm of $A^{(j)}$, thus reducing the three-axis accelerometer signal to a vector in $\mathbb{R}^M$ while preserving its energy spectral density. Transformed vibration signatures



Object Type	Abb.	Dim. (in.)	Num.	Mass (g.)
Sm. ball bearing	sbb	3/16 DIA	64	28
Ball bearing	bb	1/4 DIA	27	28
Washer	wa	3/8 DIA	78	28
Nut	nt	6-32×5/16 DIA	32	28
Screw1	s1	4-40×9/16	42	28
Screw2	s2	4-40×1 1/4	22	29
Screw3	s3	4-40×1 1/2	19	29
L-bracket	lb	9/8×5/8×3/8	11	28
Standoff	so	1×1/4 DIA	12	29
Acrylic piece	ap	1×1/4×3/2	3	28
Chocolate ball	cb	1 DIA	2	25
Rubber ball	rb	3/4 DIA	1	32

Material	Abb.	Shape	Dim. (in.)
Glass	gl	Beaker	2.5 DIA × 3.5 HT
Plastic	pl	Cup	2.5 DIA × 4.5 HT
Paper	pa	Cup	3 DIA × 4.3 HT
Cardboard	bx	Box	4.5 × 4.5 × 4

Fig. 1. Top: The 12 object types, numbered in the order listed in the table below. Middle: Object types, abbreviations and properties. Bottom: Container types, abbreviations and properties.

are now represented as vectors in $\mathbb{R}^M$. We call these the frequency vibration signatures, or just vibration signatures when the meaning is clear.

Fig. 2 shows example vibration signatures of 28 gram batches of ball bearings (27 bb) and of small ball bearings (64 sbb) in a glass beaker with $T = 1$ sec. The corresponding mean frequency signatures have a variety of high frequency and low frequency resonances. The second plot also includes the mean signature μ_e of an empty beaker. The low frequency resonances also appear in μ_e, suggesting that these are associated with the shaking mechanism, but the resonances above about 5 KHz do not, suggesting that these are associated with vibration modes of the beaker. This is in rough agreement with the vibration modes of small open glass bottles [21]. The mean signatures also show the univariate standard deviations about the mean at each frequency. These hint at the possibility of using the signatures to classify container content. However, it is important to note that we will use a multivariate approach that exploits how the class signatures spread about the respective means. Finally, as

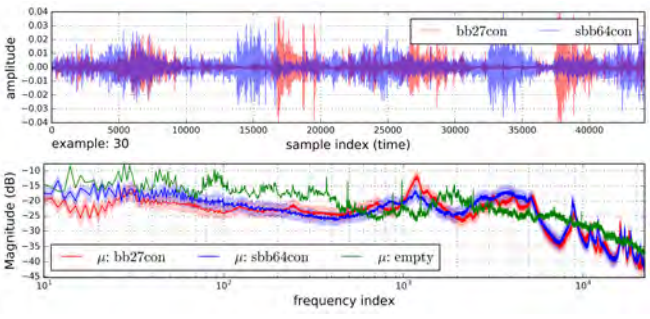


Fig. 2. Top: Superimposed one sec contact mic. signatures for 27 bb and for 64 sbb in a beaker. Bottom: The mean frequency signatures for 27 bb, 64 sbb, and an empty glass beaker on a log frequency scale. The shading indicates ±1 stand. deviation (univariate). (Best viewed in color.)

contact mic.	air mic.	accelerometer
$100.0 \pm 0\%$	$96.7 \pm 7\%$	$81.2 \pm 11\%$.

gl	pl	bx	pa
$100.0 \pm 0\%$	$100.0 \pm 0\%$	$99.2 \pm 0\%$	$97.5 \pm 5\%$.

Fig. 3. Sensor (top) and container (bottom) comparisons.

T decreases, we expect the signatures to be less informative of container content.

As observed in Fig. 2, the frequency signatures are structured with far fewer than M degrees of freedom. Principal component analysis (PCA) of training data is used to determine a projection of frequency space into $\mathbb{R}^d$ with $d \ll M$. We use $d = 4, 8, 16$ to report our results. This range was selected by applying a Gaussian Naive Bayes classifier, both before and after PCA dimensionality reduction, for a subset of pairs of objects in Fig. 1. Over $k = 5$ folds, the data was split into a training set and testing set, and we looked for the least value of $d \in [2, 1024]$ that did not result in a substantial drop in classifier test accuracy. This varied, but $d = 4 \ldots 16$ captured the relevant range for $T = 1$. Smaller values of T typically benefit by using larger values of d.

IV. Experiments & Results

Experiment 1: Sensor and Container Comparisons. We first compared performance across sensors and containers. Homogeneous 28 gram batches of washers, nuts, ball bearings, and small ball bearings were shaken in a glass beaker as each sensor captured the vibrations. One second signatures were then classified to the four object types using a multi-class linear SVM, yielding the mean accuracies in Fig. 3. Here, $\pm x$ indicates the 95% confidence interval ($\pm 1.96\times$ the standard error in the sample mean). For these experiments, the two microphones have statistically equivalent performance, while that of the accelerometer is clearly lower. It is plausible that information useful for classification is partially coded in frequencies beyond the accelerometer's bandwidth. While these initial tests did not distinguish the air and contact mic., subsequent experiments will do so.

The containers were compared using the same four 28 gram batches of objects shaken in each container as the contact microphone captured the vibrations. Pairwise linear SVM classification using the glass beaker and the plastic cup gave the greatest number of object pairs with 100% classification accuracy. We hypothesize that these containers have higher frequency vibration modes and can better exploit the 5kHz bandwidth of the sensor. However, under multi-class linear SVM classification, mean prediction accuracy only differed slightly across containers (second row Fig. 3).

In these experiments, the tested containers all yielded vibration signatures that were informative of container content. However, containers with high frequency vibration modes coupled with a higher bandwidth sensor showed advantages.

Experiment 2: Object Classification. We now examine classification of container content from vibration signatures. We first collected data by shaking homogeneous batches of the objects shown in Fig. 1 in a glass beaker. Then, we use the testing accuracy of a trained SVM to quantify how well the signatures characterized container content.

	Classification Output											
	wa	so	sbb	rb	nt	lb	cb	s1	bb	ap	s3	s2
wa	60	0	0	0	0	0	0	0	0	0	0	0
so	0	60	0	0	0	0	0	0	0	0	0	0
sbb	0	0	60	0	0	0	0	0	0	0	0	0
rb	0	0	0	57	0	0	2	0	0	1	0	0
nt	0	0	0	0	60	0	0	0	0	0	0	0
lb	0	0	0	0	0	60	0	0	0	0	0	0
cb	0	0	0	12	0	0	48	0	0	0	0	0
s1	0	0	0	0	0	0	0	60	0	0	0	0
bb	0	0	0	0	0	0	0	0	60	0	0	0
ap	0	0	0	0	0	0	0	0	0	60	0	0
s3	0	0	0	0	0	0	0	0	0	0	41	19
s2	0	0	0	0	0	0	0	0	0	0	11	49

T=1, pca d=16, Linear SVM ovr, C=0.06, cross-val. k=5

Fig. 4. Confusion matrix for multi-class classification of 12 objects.

We begin with the example of 28 gram batches of bb and sbb (Fig. 2). To test if the vibration signatures are informative of container content, we use projection to d PCA dimensions, data standardization and a linear SVM all within k-fold cross-validation. For $T = 1$, the signatures of 27 bb and 64 sbb are distinguishable with essentially 100% accuracy (contact mic.: 100%, air mic.: $99.2 \pm 4.9\%$, $d = 4$, $C = 1$, $k = 10$). To examine this in greater detail, we bring in the null hypothesis that signatures are non-informative of object class, the alternate hypothesis that signatures are informative of object class, and regard the linear SVM test accuracy as a test statistic. The density of this statistic under the null hypothesis is estimated using a label permutation test (Fig. 5). The highest values of the statistic (1, 0.99, see above) correspond to the true labels; the next highest value is 0.66. We conclude that the data provides strong evidence for the alternate hypothesis with p-value: 2×10^{-5}.

Classification of all 66 pairs of objects in Fig. 1 was tested using similar 28 gram object batches. For this dataset, the mean accuracy across all pairs is $99.0 \pm 0.6\%$. Similarly, a multi-class linear SVM across the 12 object classes yields an mean accuracy of $93.8 \pm 4\%$. The resulting confusion matrix is shown in Fig. 4. The interesting cases are when signatures are not easily distinguished. For example, chocolate ball (cb) and rubber ball (rb) signatures are often confused, as are rb and acrylic piece (ap) signatures, and screw 2 and screw 3 signatures. The rb, cb and ap classes are all examples of objects with a low count value per 28 grams, and the screws differ only by 1/4 in. in length. Aside from these cases, one sec signatures distinguished the objects with high accuracy.

Experiment 3: Object Classification with Varying Counts. We now examine more challenging problems using data

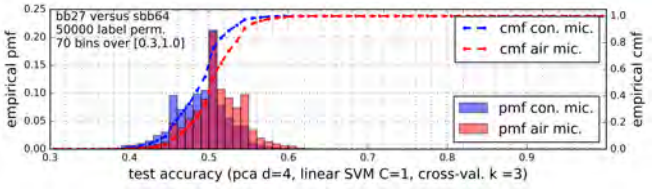


Fig. 5. Probability mass functions (pmf) and cumulative mass functions (cmf) of SVM test accuracy under the null hypothesis for the 27 bb & 64 sbb dataset. Each pmf was computed using a label permutation test.

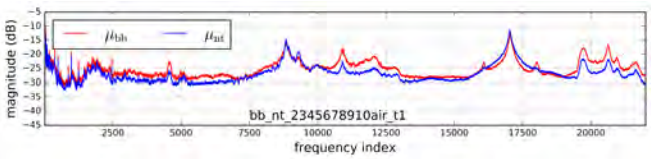


Fig. 6. Class means of the air mic. signatures for count grouping 2-10.

bb vs nt (air mic.): Classification accuracy of a left out count				
counts	pca	$T = 1$	$T = ½$	$T = ¼$
6-10	4	100.0 ± 0.0	99.3 ± 1.1	93.1 ± 5.9
	8	99.8 ± 0.7	99.6 ± 0.7	94.0 ± 5.2
	16	100.0 ± 0.0	99.8 ± 0.4	97.6 ± 2.2
4-10	4	99.6 ± 0.8	98.9 ± 1.7	92.2 ± 6.6
	8	99.8 ± 0.7	99.1 ± 0.8	93.3 ± 4.6
	16	99.8 ± 0.7	99.6 ± 0.7	97.1 ± 3.0
2-10	4	99.0 ± 4.1	98.2 ± 3.8	92.2 ± 6.9
	8	99.8 ± 0.7	98.7 ± 3.0	92.1 ± 5.9
	16	99.7 ± 1.1	99.3 ± 1.8	94.8 ± 8.0
bb vs nt (con. mic.): classification accuracy of a left out count				
counts	pca	$T = 1$	$T = ½$	$T = ¼$
6-10	4	99.0 ± 1.9	99.2 ± 2.1	91.2 ± 9.2
	8	99.7 ± 0.8	99.2 ± 2.6	93.2 ± 8.5
	16	99.8 ± 0.7	99.1 ± 2.2	97.2 ± 3.9
4-10	4	94.4 ± 9.5	82.8 ± 31.4	81.1 ± 25.3
	8	96.9 ± 5.8	87.3 ± 30.8	77.1 ± 25.8
	16	97.6 ± 4.3	90.2 ± 26.5	84.7 ± 18.5
2-10	4	90.7 ± 15.3	80.7 ± 26.8	71.4 ± 21.5
	8	92.3 ± 13.7	85.4 ± 20.9	72.8 ± 18.9
	16	94.4 ± 10.3	85.0 ± 24.0	80.5 ± 13.4

Linear SVM, C=1, cross-validation k=10

Fig. 7. bb versus nt: Accuracy of label prediction for a left out count.

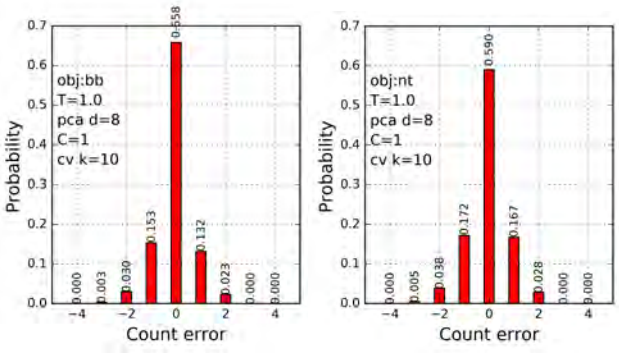
gathered by shaking x nuts (nt) and separately x ball bearings (bb), for $x = 1, \ldots, 10$. This data is then grouped across multiple count values for each object as follows: counts of 6 to 10, 4 to 10, and 2 to 10. In each grouping, the signatures range over multiple count values, and the data for each value is obtained from a separate 60 sec run of the shaker. For $T = 1$, this yields 600, 840, and 1080 labeled examples, respectively. Fig. 6 shows the 2-10 grouping class means.

For $T = 1$, ½, ¼ and projection dimensions $d = 4, 8, 16$, a linear SVM is used to classify the grouped signatures between the classes bb and nt. We tested classification accuracy by holding out one count value (for both objects), training on the rest of the data and then testing on the held out data. For example, for the grouping 6-10, we hold out all 6 count data, train on the data for 7 to 10 counts, then predict the labels for the 6 count data. This is repeated for each count value and the results are averaged. We call this leave-one-count-out cross-validation. This is a more demanding test metric than k-fold cross-validation, since test signatures have a count value not seen during training. This is reflected by slightly worse performance than k-fold cross-validation in the more challenging tests (e.g. the grouping 2-10). Classifying $T = 1$, ½ sec air mic. signatures with $d = 16$ yields mean accuracies hovering above 99% for all groupings (Fig. 7). The contact mic. yields very competitive results for the 6-10 grouping for all three values of T, but its performance begins to drop on the more challenging groupings 4-10 and 2-10. Examination of Fig. 6 suggests that the air mic. may be better at capturing high frequency data around the beaker resonances. As expected, accuracy decreases as T decreases. As the PCA dimension d increases, accuracy can initially increase (assuming there is headroom to do so) but this effect can quickly saturate.

	Classification Output: counting bb									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	98	2	0	0	0	0	0	0	0	0
2	2	85	12	2	0	0	0	0	0	0
3	0	18	58	23	0	0	0	0	0	0
4	2	5	13	73	7	0	0	0	0	0
5	0	0	2	12	73	13	0	0	0	0
6	0	0	0	0	13	53	18	15	0	0
7	0	0	0	0	0	13	67	17	3	0
8	0	0	0	0	0	8	23	50	15	3
9	0	0	0	0	0	0	8	22	45	25
10	0	0	0	0	0	0	2	7	37	55

	Classification Output: counting nt									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	95	5	0	0	0	0	0	0	0	0
2	3	85	12	0	0	0	0	0	0	0
3	0	12	55	25	8	0	0	0	0	0
4	0	0	25	43	27	5	0	0	0	0
5	0	0	2	22	52	22	3	0	0	0
6	0	0	0	3	7	72	15	3	0	0
7	0	0	0	0	5	33	37	20	5	0
8	0	0	0	0	0	7	23	50	17	3
9	0	0	0	0	0	2	8	23	42	25
10	0	0	0	0	0	0	3	13	23	60

T=1, pca d=8, Linear SVM, C=1, cross-validation k=10.



	$T = 1$		$T = ½$	
	bb	nt	bb	nt
Exact:	65.8%	59%	51.3%	45.9%
Within ± 1:	94.3%	92.8%	87.1%	86.3%
Within ± 2:	99.7%	99.5%	98.4%	96.7%

Fig. 8. Top: Confusion matrices for counting bb and nt items. Entries are percentages rounded to the nearest integer. Bottom: Empirical error pmf for bb (left) and nt (right), and a summary of accuracies.

Experiment 4: Counting Items. The bb and nt data in Exp. 3 is now used to explore how well the signatures determine the number of objects in the container. We use signatures with $T = 1$, ½ sec, giving a total of 600 and 1,200 labelled examples, respectively. For classification into 10 count classes, we use projection to d PCA dimensions, data standardization and a multi-class linear SVM all within k-fold cross-validation. The resulting confusion matrices and the empirical error distributions are shown in Fig. 8. Chance accuracy is 10%. The mean accuracy for exact counting is well above chance, but not impressive. However, counting to within a tolerance of ± 2 is possible with mean accuracy $\geq$ 99%. The results also suggest that signatures are less informative of item count for high count values. Nevertheless, signatures can still answer questions that respect this constraint, e.g., is the beaker nearly empty or not ($\{1, 2, 3\}$ vs $\{8, 9, 10\}$)? In this problem, it is critical to classify the 1,2,3 and 8,9,10 counts correctly, but errors for 4,5,6,7 counts

are unimportant. For the bb data, a linear SVM as in Fig. 8 answers this question with 100% accuracy.

Experiment 5: Classifying Transient Data. The prior experiments have used stationary signatures, i.e. vibration signatures recorded after initial transients decay. In this experiment, we test if a classifier trained on one sec stationary signatures can correctly classify transient signatures captured in 2 sec bursts (capturing both initial and final transients). Separate 28 gram batches of nuts, washers, ball bearings, and small ball bearings were shaken 10 times for a 2 sec burst. Since $T = 2$ sec for these signatures, we subsampled the DFT data to maintain the same-length feature vectors as before. A multi-class linear SVM trained on one sec stationary vibration signatures classified these 10 transient signatures with a mean accuracy of 97.5%. This is slightly below the 100% accuracy for one sec stationary data.

Experiment 6: Unsupervised Categorization. A key ingredient in exploratory robotic perception is the ability to relate new object data to previously seen data. One way to do so is through data clustering. For example, objects can be clustered through the lens of training vibration signatures. Moreover, if the training data are tagged with attributes of the corresponding objects, the attributes are transferred to the clusters to which the vibration signatures are assigned.

We examine this idea for the vibration signatures of 28 gram batches of objects. In the training phase, PCA projection into $\mathbb{R}^d$ is determined by the aggregate training data. K-means clustering is then applied to the projected signatures. The learned PCA projection is applied to the held-out data, and each projected signature is assigned to its nearest cluster (Fig. 9). In the example shown, the data from 9 objects are used to learn the clustering and the data for 3 held-out objects are then mapped to the clustering.

At $k = 5$, the clusters *begin* to contain objects with shared characteristics, e.g., shapes, dimensions, counts. The objects in cluster 1 are similar in scale and each has a circular symmetry, those in cluster 4 share a hexagonal contour, and the screws in cluster 5 only differ in length by ¼ in. Cluster 2 contains objects with the least items per 28 grams (Fig. 1).

The held-out data are assigned into natural clusters for $k = 2, \ldots, 9$. Observe that the acrylic piece is assigned between clusters 2, the low item count cluster, and 3. In cluster 3, the angularity of the l-brackets and acrylic pieces might render their vibration signatures similar, and indeed they remain in the same cluster at $k = 9$. The only odd clustering at $k = 9$ is the grouping of ball bearings with washers instead of small ball bearings. When viewed in 3 dimensional PCA space, the three categories are distinct and equidistantly separated.

At $k = 9$, the rubber ball (rb) is the only object with an ambiguous signature clustering. Because of the beaker's cylindrical shape and its slightly convex bottom, the single rubber ball often circled inside the beaker rather than hitting the beaker wall. As a result, about half of the rb vibration signatures cluster near those of the chocolate ball, while the rest spread away from that region. This explains its clustering ambiguity. The effect is visualized in Fig. 9 by projecting the rb and cb signatures onto the first three principal components.

k	Clusters
2	[wa,sbb,bb], [so,rb,nt,lb,cb,s1,s2,s3,ap]
3	[wa,sbb,bb], [rb,cb,ap], [so,rb,nt,lb,s1,s2,s3,ap]
4	[wa,sbb,bb], [rb,cb,ap], [so,rb,lb,s1,ap], [nt,s2,s3]
5	[wa,sbb,bb], [rb,cb,ap], [rb,lb,s1,ap], [so,nt], [s2,s3]
⋮	⋮
9	[wa,bb], [sbb], [rb], [rb,cb], [lb,ap], [s1], [so], [nt], [s2,s3]

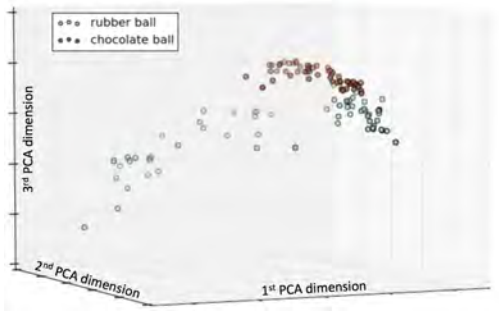


Fig. 9. Top: The vibration signatures of 9 objects separated into k clusters using K-means clustering. The data of held-out objects (in red) are then mapped to the clusters. See Fig. 1 for object name abbreviations. Bottom: Rubber and chocolate ball signatures mapped to first three PCA dimensions.

V. Discussion and Conclusion

We have examined a simple tactile sensor system for identifying contents shaken in a container. Simple, standard machine learning methods successfully extracted nuanced information from the measured vibration signatures. In addition to classifying known objects, say in a structured setting, this also permits new objects to be tagged with the attributes of known training data, or to be distinguished from (previously seen) objects with a small error margin. Our novel contribution involves the examination of challenging tasks in this context. For example, we have shown that vibrations can be used to count nuts from 1-10 in a glass beaker within an error of ± 2 with 99% accuracy, to classify objects from a short 2 sec burst of shaking at a high accuracy, and to distinguish ball bearings from nuts across multiple counts. We regard these as preliminary proof-of-concept results indicating considerable potential for exploiting tactile vibration sensing to aid robotic perception. In practice, vibration sensing can be fused with complementary sensing modalities to enhance the overall robustness of sensory perception.

An enlightening trade-off is observed by comparing the performances of the accelerometer, contact mic., and the studio quality air mic. The accelerometer is the least expensive and has the smallest bandwidth. In our tests, it gave classification performance well above chance but lower than the other sensors. The contact mic. is two orders of magnitude less expensive than the air mic. and is more robust to environmental noise and poor acoustical settings (e.g. in space). However, the air mic. has a much higher bandwidth (this is reflected in higher accuracies for more difficult tasks) and does not rely on colocation with the container. The choice of sensor therefore depends on both the circumstances of use and the desired performance.

We foresee applications of this work in structured settings (e.g., stores, workshops, warehouses, homes) where it is feasible for service robots to extensively explore the vibration signatures likely to be encountered. To that end, a significant expansion of the vocabulary of vibration signatures will allow a more complete evaluation of the utility of vibration sensing. However, vibration sensing may also be useful in less structured settings (e.g., disaster relief, construction, mars rover) where previously unexamined objects may be encountered. We have provided some thoughts on this in our discussion of signature clustering, but this is largely an area open for further investigation.

ACKNOWLEDGEMENTS

The first author acknowledges the support of senior thesis funds from the Dept. of Electrical Engineering and School of Engineering and Applied Science at Princeton University, and thanks Prof. Katherine Kuchenbecker, David Radcliff, Anthony Sigillito, Matthew Matl, Hossein Valavi, Prof. Ken Goldberg and the AUTOLab at U.C. Berkeley for their valuable feedback, insight, and support.

APPENDIX: MACHINE LEARNING METHODS

This section briefly describes the machine learning methods employed. We used the implementations in Python 2.7's Scikit-learn toolbox [22]. The data $\{(x_i, y_i)\}_{i=1}^n$ consists of examples in $x_i \in \mathbb{R}^d$ and corresponding labels $y_i \in L$, for a finite set L. A learning algorithm is said to be unsupervised if it does not use the labels and supervised otherwise.

Training and Testing Data: Training data is used to learn a classifier/model and testing data is used to estimate its post training performance. It is crucial that testing data is not used during the learning phase. This is done by splitting the available data into a training set and a disjoint testing set.

k-Fold Cross-Validation: is used for creating multiple training and testing sets from a finite set of labeled data. The labeled data is divided into k folds; (k-1) folds are used as training data and the left out fold as testing data. The folds are stratified if each has a balanced label set. PCA projection and data standardization can be applied to the training data prior to classifier learning. The same fixed transformations are then applied to the testing data. This process is iterated over the left out fold and the k results are used to estimate mean classification accuracy and 95% confidence bounds.

Gaussian Naive Bayes Classifier (GNB): For each label j, GNB finds the empirical mean μ_j, and the best fit spherical covariance $\sigma_j^2 I$ of the data with label j. Given the label, the data is treated as being generated independently according to a multivariate Gaussian density with these parameters and testing data is assigned the class under which it is most likely.

Support Vector Machine (SVM): For binary classification, the linear SVM solves a quadratic optimization problem using the training data to learn a hyperplane that best "separates" the training data into two classes. It has one regularization parameter $C > 0$ that needs to be selected. The SVM can also be used to perform multi-class classification, e.g., by using one-versus-rest or one-versus-one approaches.

K-Means: is an unsupervised clustering algorithm based on Euclidean distance. For a given integer k, the algorithm finds k cluster centers μ_i and an assignment of each data point to one μ_i such that data points are assigned to the closest cluster center, and μ_i is the mean of the data points assigned to it. Because of multiple local minima, K-means is applied with several random initializations and the best result is selected.

REFERENCES

[1] J. M. Romano *et al.*, "Human-inspired robotic grasp control with tactile sensing," *IEEE Trans. on Robotics*, vol. 27, no. 6, pp. 1067–1079, Dec 2011.

[2] A. Garg *et al.*, "Tumor localization using automated palpation with gaussian process adaptive sampling," 2016, preprint submitted to 12th Conference on Automation Science and Engineering.

[3] S. McKinley *et al.*, "A disposable haptic palpation probe for locating subcutaneous blood vessels in robot-assisted minimally invasive surgery," *IEEE CASE*, 2015.

[4] J. M. Romano and K. J. Kuchenbecker, "Methods for robotic tool-mediated haptic surface recognition," in *IEEE Haptics Symposium*, 2014, pp. 49–56.

[5] M. Strese *et al.*, "Surface classification using acceleration signals recorded during human freehand movement," in *IEEE World Haptics Conference*, 2015, pp. 214–219.

[6] J. A. Fishel and G. E. Loeb, "Bayesian exploration for intelligent identification of textures," *Front Neurorobot*, vol. 6, no. 4, 2012.

[7] J. Hoelscher *et al.*, "Evaluation of tactile feature extraction for interactive object recognition," in *IEEE 15th Int. Conf. Humanoid Robots*, 2015, pp. 310–317.

[8] M. Kaboli *et al.*, "In-hand object recognition via texture properties with robotic hands, artificial skin, and novel tactile descriptors," in *IEEE 15th Int. Conf. Humanoid Robots*, 2015, pp. 1155–1160.

[9] I. Abu-Mahfouz, "Drilling wear detection and classification using vibration signals and artificial neural network," *International Journal of Machine Tools and Manufacture*, vol. 43, no. 7, pp. 707–720, 2003.

[10] W. Ostachowicz and M. Krawczuk, "Analysis of the effect of cracks on the natural frequencies of a cantilever beam," *Journal of sound and vibration*, vol. 150, no. 2, pp. 191–201, 1991.

[11] F. L. di Scalea and J. McNamara, "Measuring high-frequency wave propagation in railroad tracks by joint time–frequency analysis," *Journal of Sound and Vibration*, vol. 273, no. 3, pp. 637–651, 2004.

[12] S. Finnveden, "Spectral finite element analysis of the vibration of straight fluid-filled pipes with flanges," *Journal of Sound and Vibration*, vol. 199, no. 1, pp. 125–154, 1997.

[13] W. A. Wilkinson and M. Cox, "Discrete wavelet analysis of power system transients," *Power Systems, IEEE Transactions on*, vol. 11, no. 4, pp. 2038–2044, 1996.

[14] K. D. Marshall, "Modal analysis of a violin," *The Journal of the Acoustical Society of America*, vol. 77, no. 2, pp. 695–709, 1985.

[15] C. A. Brooks and K. Iagnemma, "Vibration-based terrain classification for planetary exploration rovers," *IEEE Trans. on Robotics*, vol. 21, no. 6, pp. 1185–1191, 2005.

[16] J. Libby and A. J. Stentz, "Using sound to classify vehicle-terrain interactions in outdoor environments," in *IEEE Int. Conf. Robotics and Automation*, 2012, pp. 3559–3566.

[17] S. Takamuku *et al.*, "Shaking eases object category acquisition: Experiments with a robot arm," in *Proc. Seventh Int. Conf. on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems.* Lund University Cognitive Studies, 135., 2007.

[18] J. Sinapov *et al.*, "Learning relational object categories using behavioral exploration and multimodal perception," in *Int. Conf. Robotics and Automation*, 2014, pp. 5691–5698.

[19] STMicroelectronics, "LIS344ALH," Datasheet, 2008.

[20] N. Landin *et al.*, "Dimensional reduction of high-frequency accelerations for haptic rendering," in *Haptics: Generating and Perceiving Tangible Sensations*. Springer, 2010, pp. 79–86.

[21] D. A. Russell, "Acoustics and vibration animations," http://www.acs.psu.edu/drussell/demos.html, 2011.

[22] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.