Soft Tactile Sensors for Robot Grippers Using Acoustic Sensing

Kevin Xu*[1] and Justin Chan[1]

***Abstract*— We present a low-cost, soft tactile sensor using common, easily sourced materials that can be integrated with existing robotic gripper systems without requiring complex fabrication techniques or expensive components. Our approach includes two designs: a flexible linear sensor constructed from a rubber tube and a planar sensor made with a rubber membrane stretched over an enclosure. Both sensors contain an embedded speaker and microphones that leverage active acoustic sensing to map the unique acoustic resonant response of the cavity's structure to deformations that occur when the robotic gripper is grasping an object. Experimental results demonstrate that, using a support vector machine, the linear sensor achieves contact point estimation with an RMSE of 6 mm, while the planar sensor achieves an RMSE of 0.57–0.62 mm. Additionally, the planar sensor classifies six objects with an accuracy of 97.7%. These results demonstrate the potential for active acoustics to be an accessible method for enabling tactile sensing capabilities for robotic systems.**

I. Introduction

Touch as a sensory modality provides important information about an object's material properties, such as shape and contact location. Here, we ask the question: *can we create a low-cost, soft tactile sensor using common materials that can be attached to robotic grippers?* The integration of soft tactile sensors onto robotic grippers and other end-effectors enables more precise object manipulation, handling of irregular objects, and safer human-robot interactions.

We present the design of a soft, flexible tactile sensor made from common, low-cost materials that can be retrofitted onto existing robotic grippers to provide tactile sensing capabilities. While state-of-the-art tactile sensors typically require expensive materials or advanced fabrication techniques for high-fidelity measurements, these factors can limit their accessibility and scalability. Our work is focused on an affordable sensor design which can enable researchers from various disciplines and non-specialists to integrate tactile sensing into robotic systems. Our design uses readily available materials, making it reproducible in low-resource environments, and operates without vibrating, tapping, or scratching objects, making it appropriate for use with fragile and delicate items.

Our sensor design leverages active acoustic sensing by using a speaker to emit a wideband pulse into the sensor's interior while a microphone captures the acoustic response – effectively creating a fingerprint that uniquely characterizes the internal geometry of the sensor. When the soft sensor makes contact with an object, its geometry changes, altering the acoustic multipath profile within the sensor and measurably changing the acoustic response. By developing a library of acoustic responses corresponding to different sensor perturbations, we show that it is possible to perform contact point estimation to distinguish between depressions at different points along the sensor, as well as classify objects with distinct geometric profiles.

We show two different designs: (1) A linear, tubular sensor that produces resonant peaks corresponding to standing waves within the tube's geometry. Perturbations to the tube attenuates the signal and creates mode coupling between the resonances of the speaker and microphone, resulting in shifts or elimination of resonant frequencies. (2) A planar

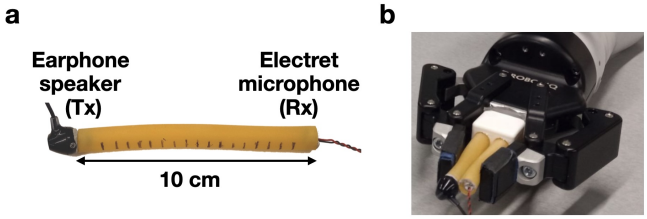


Fig. 1: **Linear tactile sensor design. (a)** Speaker and microphone are coupled to opposing ends of the rubber tube. Surface markings used for contact location estimation evaluations are separated by 5 mm. **(b)** Linear tactile sensor attached to existing robotic gripper arm (Robotiq 2F-85 gripper end-effector [1]) to enable sensing of different objects.

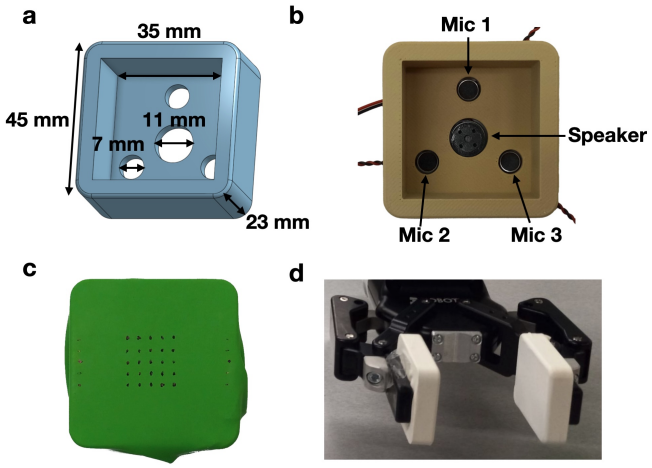


Fig. 2: **Planar tactile sensor design. (a)** CAD drawing with dimensions. **(b)** Location of speaker and microphones. **(c)** Rubber membrane interface for tactile sensing marked with a 5×5 grid to evaluate contact location estimation. **(d)** Planar sensor attached to robotic gripper arm.

* Funded by the National Science Foundation Graduate Research Fellowship Program, grant number: DGE2140739

[1] Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA {kx2,justinc3}@andrew.cmu.edu

sensor with a flexible rubber membrane stretched over a cavity containing a speaker and three microphones, where pressure applied to the surface alters the resonant response at the multiple microphones. By combining the correlated information across multiple microphones, we are able to more reliably perform tactile sensing tasks. Our experiments demonstrate that the linear sensor design can perform contact point estimation with a resolution of 6 mm, and our planar sensor can classify between six different grasped objects with an accuracy of 97.7%. We conducted benchmarks examining the effect of different depression depths, ambient background noise levels, multiple microphones, classifier types, and transmission signals.

The remainder of this paper is organized as follows. A review of related work on tactile and acoustic sensing is given in Sec. II; an overview of the tactile sensor design is presented in Sec. III; the experimental results are described in Sec IV; and discussion and future work are discussed in Sec. V.

II. Related work

A. Tactile sensing

Tactile sensing systems provide rich feedback to a robot to understand the physical properties of an object. Vision-based techniques have used a soft elastometer surface as a touch-sensing interface, structured light to illuminate the interior base of the membrane, and a camera to measure the deformations when the sensor makes contact with an object in high resolution [2]–[8]. Cameras have also been used to directly image the grasped object, and combine RGB images with tactile cues collected from other sensing modalities [9]. Torque sensors [10], liquid metal strain sensors [11], [12], optical waveguides [13], magnetic particles and signals [14], and tactile skins made of fabric [15], silicone [16], and wire-drawing coating [17] have also been used for contact point estimation and object classification. While many of these tactile sensors can obtain high-fidelity material information about an object, they typically require high-cost or specialized domain knowledge to fabricate and reproduce, and may have to be co-designed to be compatible with the robotic end-effector. In this work, our focus is on a low-cost, easy-to-reproduce method for tactile sensing which can be retrofitted onto existing robotic grippers using a flexible sensor design that can be created using commonly sourced material.

B. Acoustic sensing

Acoustic sensing systems for tactile sensors have the advantage of relying on low-cost components, specifically a speaker and a microphone. Acoustic sensing techniques are often fused together with vision and haptic sensing modalities to jointly solve complex manipulation tasks [8], [18]. Acoustic sensing systems can be categorized into active and passive sensing systems. Active sensing systems emit sound from a speaker and record the acoustic response at a microphone. Passive sensing systems rely solely on the acoustic recordings at the microphone caused by the contact event between the sensor and an object. Prior works have

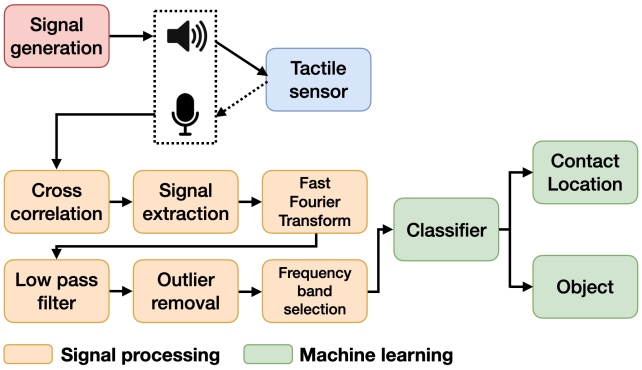


Fig. 3: **Tactile sensing pipeline for contact point estimation and object classification.**

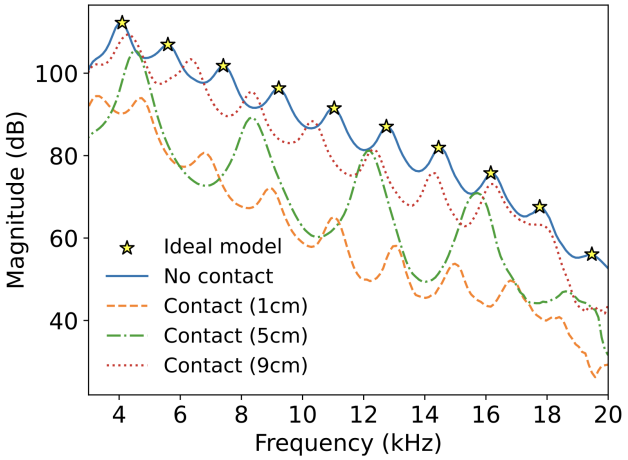
leveraged active acoustic sensing as a length sensor for a pneumatic actuator using the resonance characteristics [19], to estimate contact force and location [20], and to decode braille displays [21]. Active acoustic sensing systems have also placed multiple microphones along the length of an actuator to estimate its shape and perform robotic proprioception, by collecting a dataset of acoustic responses for different actuator shapes [22], [23]. Acoustic sensors have also been incorporated into robotic arms for collision and proximity detection [24]. In addition to speakers, bone-conduction actuators have also been used to impart vibrations onto an object to detect object state, such as material or liquid type [25]. Passive sensing systems have leveraged existing sound sources such as the noise of an air pump to perform contact point estimation along an air tube with a contact microphone [26]. The acoustic signature produced during a contact event has also been used to learn object properties and improve the performance of robot manipulation tasks [27], [28]. We build on top of these prior works by exploring whether we can leverage a single and multi-microphone design with a low-cost flexible tactile sensor in a linear and planar form factor that can be easily attached to existing robotic grippers tactile sensing tasks.

III. System overview

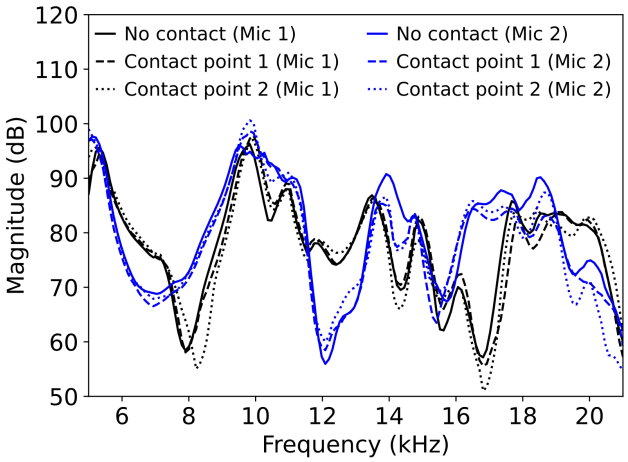
A. Sensor design

Linear Tactile Sensor Design. Our linear sensor design is comprised of a flexible rubber tube (Home Depot, Natural Latex Tubing [29]) that is of length 10 cm, with inner diameter of 0.635 cm, and outer diameter of 0.953 cm (Fig. 1). We coupled a speaker element to one end and a microphone to the other end. Specifically, we used an off-the-shelf earphone (Sony MDR-EX15AP [30]) as the speaker element, and an electret condenser microphone (Same Sky CMC-6027-32L100 [31]) as the microphone element. To measure the acoustic response to different perturbations of the sensor, we transmit 50 ms Gaussian pulses with frequency 20 Hz at a sound level of 54 dBA.

Planar Tactile Sensor Design. The planar sensor is comprised of a rubber membrane from a balloon stretched tightly across a 3D printed hollowed-out square prism with four holes at the base (Fig. 2). No additional adhesive is needed



(a) Linear tactile sensor



(b) Planar tactile sensor

Fig. 4: **Resonant acoustic response of the tactile sensors. (a)** Response from the linear tactile sensor with and without contact at different points. Idealized resonant frequency values, based on a closed-form model for an unperturbed sensor, are marked. **(b)** Response from the planar tactile sensor at two spatially separated microphones, with and without tactile contact at different points.

for the rubber membrane to stay in place, and the membrane stayed secure throughout evaluation of the sensor's performance. The speaker element is coupled to the center hole, while three microphone elements are coupled to the top, left, and right holes. The sensor's base has dimensions of 4.5×4.5 cm, and the top opening has dimensions of 3.5×3.5 cm. We select a magnet speaker with a dynamic driver (Fielect Magnetic Speaker [32]) due to its ability to transmit at a louder volume than the earphones; in our system, we select a sound level of 64 dBA. We used the same electret condenser microphones as the linear sensor. We also transmit pulses at a frequency of 20 Hz.

B. Acoustic model

Our linear sensor design made of rubber in an unpressed state can be approximated as a closed tube. When a speaker transmits sound at one end of the tube, the sound reflects off the other end of the tube, interferes with the transmitted wave, and creates standing waves within it. These standing waves produce an acoustic resonance pattern due to interaction between the speaker, the medium (air), and the walls of the cavity. This resonance response can be observed using a microphone that is located at the other end of the tube, where a pressure antinode is formed; this is the location along the tube where maximum pressure variation is observed due to constructive interference between the incident and reflected waves.

Linear sensor

Component	Cost (USD)
Earbud [30]	8.19
Microphone [31]	1.22
Rubber tube [29]	0.72
Total	**10.13**

Planar sensor

Component	Cost (USD)
Speaker [32]	2.50
Microphone [31]	3.66
Audio adapters [33]	15.98
Balloon	0.20
Total	**22.34**

TABLE I: Itemized hardware cost.

The resonance frequencies f_n where n is the harmonic number $(1, 2, 3, \ldots)$ are periodically spaced frequencies where there is a peak in the acoustic response, and are related to the length of the tube L. Let v be the speed of sound in the medium which is 343 m/s for air, and f_{offset} is a system-dependent frequency offset. The need for an offset is a correction factor that accounts for the tube diameter, tube material, and viscous and thermal losses at the soft tube walls. The resonant frequency f_n can be denoted by Eq. (1):

$$f_n = \frac{(2n-1)v}{4L} + f_{offset} \tag{1}$$

Fig. 4a shows the measured resonant response of the 10 cm linear sensor after transmitting a Gaussian pulse of duration 15.8 ms. The measured response is shown when there is no contact, and the ideal resonant frequencies are computed with the closed-form equation using $f_{offset} = -211$ Hz, which minimizes the mean error between the ideal and measured resonant frequencies (mean error: 0.1 Hz). When contact pressure is applied at different locations of the tube, this results in a shift or elimination of resonant frequencies, as well as an attenuation of acoustic energy within the tube. This is because the speaker and the cavity have independent resonance modes, and the application of contact pressure results in *mode coupling [34]* or cross interaction between the resonances resulting in frequency shifts. *The key point to note is that the level of attenuation and the resonant frequencies within the acoustic response are uniquely related to contact point location.*

For the planar sensor, we can observe in Fig. 4b that the resonances do not occur at periodic frequencies. This is because the resonance frequencies for rectangular objects denoted $f_{l,m,n}$ are defined by three independent non-negative mode indices l, m, n and three dimensions of the sensor L_x, L_y, L_z as in Eq. (2):

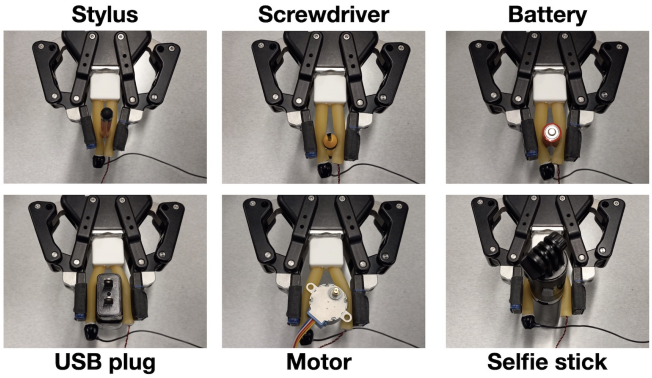


Fig. 5: **Object classification experimental setup.** The robotic gripper was tasked with grasping six objects of varied geometric profiles using the linear tactile sensor.

$$f_{l,m,n} = \frac{v}{2}\sqrt{(\frac{l}{L_x})^2 + (\frac{m}{L_y})^2 + (\frac{n}{L_z})^2} \tag{2}$$

This can result in clustering or irregular gaps in the frequencies, and in the case where the sensor was a cuboid ($L_x = L_y = L_z$), different mode combinations can yield the same resonant frequency and lead to degenerate modes. *The main observations in Fig. 4 are: (1) contact with the rubber membrane will create measurable attenuations to the resonant frequency response; (2) contact at different points produce different resonant responses; and (3) the measured response differs between microphones as the multipath characteristics change at different points in space. By leveraging data across multiple microphones, we can uniquely learn the mapping between the acoustic response and the contact point on the membrane.*

C. Signal processing

In our system (Fig. 3), we transmit a Gaussian pulse with a sampling rate of 48 kHz and duration of 15.8 ms to characterize the frequency response of the sensors. There is a 50 ms guard band in between pulses, where no signal is transmitted; this allows reflections and reverberations from the previous pulse to decay, so as not to interfere with measurements of the next pulse. For each measurement, we apply cross-correlation to the received signal and extract the acoustic measurement for candidate peaks that exceed an amplitude threshold. For each candidate measurement, we perform a 450-point fast-Fourier transform (FFT) to transform the time-domain signal into the frequency-domain. We then smooth the signal using a moving average filter with window size 10 to remove the effects of high-frequency noise. We extract the frequency response within the range of 2.5–20 kHz for the linear sensor and 5–21 kHz for the planar sensor; we empirically observed these ranges contained the most amount of variability upon applying pressure at various points along the sensor. The amplitude values of the acoustic response are converted into decibels. This resulted in a feature vector of size of 165 and 150 points for the linear and planar sensor respectively. However, since the cross-correlation can capture spurious peaks from noise not resulting from the transmitted pulse, we implement heuristics to remove outlier outputs. Specifically, if the average amplitude across the acoustic response captured in the feature vector is less than 50 dB, and the standard deviation is less than 3 dB, the candidate response is discarded.

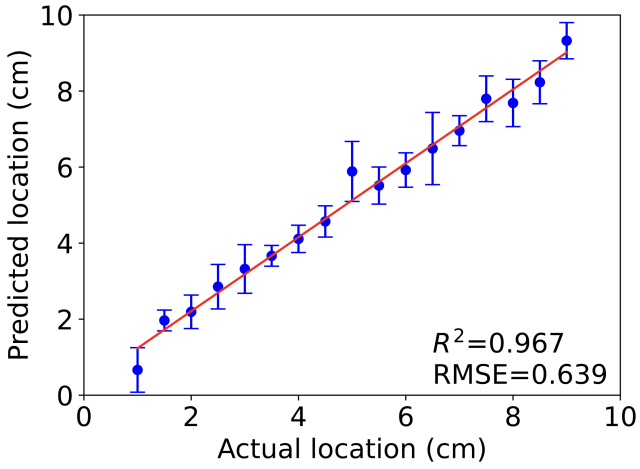


Fig. 6: **Spatial accuracy of linear tactile sensor.** Linear regression performed on the acoustic responses measured along the full length of the tube shows an RMSE of 6.4 mm.

IV. Feasibility Results

Experimental setup. The earphone and microphone were connected to a Samsung Galaxy S9 smartphone via a screw terminal connector cable. Using a custom smartphone application, we transmitted the Gaussian pulses at a frequency of 20 Hz, while simultaneously recording from the electret condenser microphone. The sampling rate for the speaker and microphone was set to 48 kHz. In all following experiments, one experimental trial refers to pressing the sensor, removing the pressure, and extracting 10 acoustic responses while the sensor is pressed. We note that in the evaluations, there is no overlap in trial data between the training and testing set.

A. Linear tactile sensor experiments

Contact location estimation. We discretize the surface of the sensor into 17 points with an interval of 0.5 cm ranging from 1 to 9 cm along the length (Fig. 1). We applied pressure at each contact point by using a finger to depress the tube fully, and we held the probe in place for 7 s before releasing it. In our dataset, we collected 10 contact trials at each of the 17 contact points, for a total of 170 trials. To determine if our system is able to classify between the acoustic responses generated across the 17 locations, we leverage a support vector machine (SVM) with a linear kernel and perform five-fold cross-validation on the acoustic frequency responses to estimate the contact location. We obtain an accuracy of 97.82 $\pm$ 0.02%. Additionally, we apply a linear regression model on the acoustic response data, using an 80-20 train-test split. As shown in Fig. 6, the linear regression model produced a coefficient of determination (R^2) of 0.967, with an RMSE of 0.639 cm.

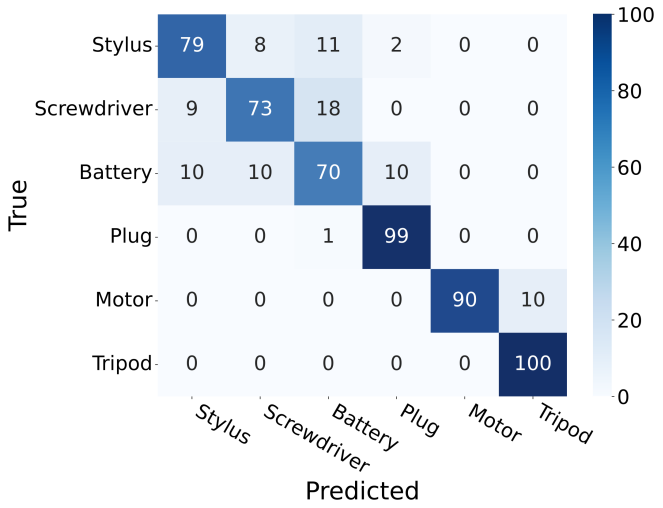


Fig. 7: **Confusion matrix for classifying 6 objects with linear tactile sensor.**

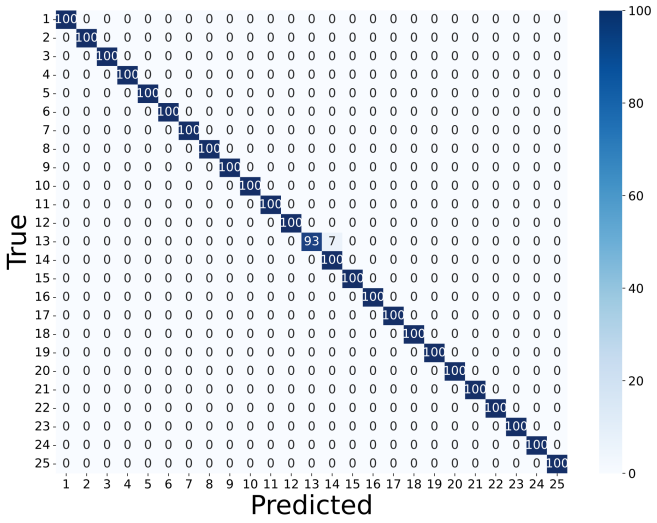


Fig. 8: **Confusion matrix for contact point location estimation along the 5×5 grid on the planar tactile sensor.**

Object classification. To evaluate the ability of the linear sensor to classify between different objects, we attached the sensor to a robotic arm gripper (Robotiq 2F-85 gripper end-effector [1]). We first thread the sensor through a 3D printed connector so it forms a U-shape. The connector prevents the tubing from compressing and pinching at the midpoint, which would prevent sound from being transmitted through. The connector is then secured to the gripping center, and the tubing is attached to the left and right arms of the gripper.

We selected six objects (tablet stylus, screwdriver, AAA battery, USB wall charger, DC motor, selfie stick) which could be gripped by the robotic gripper, and which had different side profiles (Fig. 5). We instructed the robot gripper to grasp the object, measured the acoustic response, and instructed the gripper to ungrasp the object. This was repeated with 10 trials for each object, as well as 10 frequency responses extracted for each trial for each of the 6 objects, resulting in a dataset of 600 acoustic responses.

We trained an SVM with a linear kernel using five-fold cross validation, which produced an accuracy of 85.2 $\pm$ 7.02% as shown in the confusion matrix in Fig. 7. We note that misclassifications are common among the stylus, screwdriver, and battery, which have similar side profiles due to all being cylindrical objects with a similar diameter (in contrast to the shapes of the other objects). These results suggest that the sensor could serve as a coarse-grained object classifier for objects with sufficiently distinct geometries.

B. Planar tactile sensor experiments

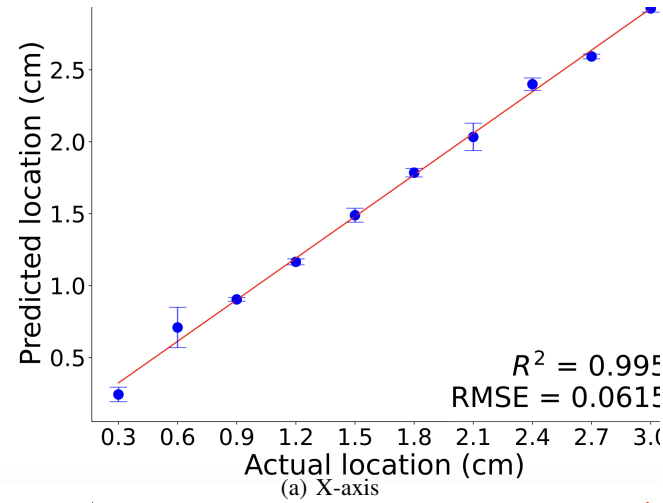
Contact location estimation. Here, we discretize the rubber membrane of the planar sensor into a 5×5 grid with an X and Y resolution of 3 mm between points, and evaluate the system's ability to classify between different points. We perform 10 trials at each of the 25 points for a total of 250 measurements. Unlike the linear sensor, for the planar sensor, the acoustic response is extracted across 3 microphones, and concatenated into a single array for further processing. The data was classified with 5-fold cross-validation, using a SVM with linear kernel, with an average accuracy of 99.7 $\pm$ 0.6% (Fig. 8). Next, we perform a linear regression on acoustic responses collected along 10 points on the x-axis and 10 points on y-axis (Fig. 9). When applying linear regression to the x and y-axis, the R^2 coefficient was 0.992 and 0.985 respectively, and the RMSE is 0.62 mm and 0.57 mm respectively.

TABLE II: Planar sensor contact location estimation performance when tested on an unseen microphne.

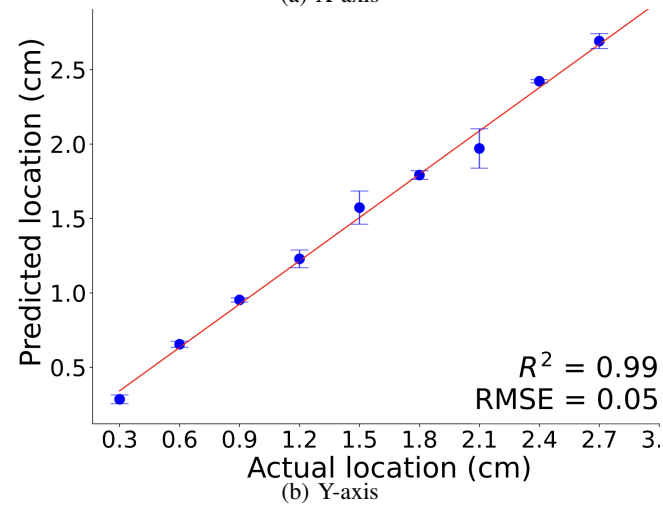
Train microphone	Test microphone	Accuracy (%)
Mic 1	Mic 2, 3	99.4
Mic 2	Mic 1, 3	96.9
Mic 3	Mic 1, 2	94.7
Mic 1, 2	Mic 3	99.2
Mic 2, 3	Mic 1	99.0
Mic 1, 3	Mic 2	98.3

We analyze the effect of training our model using data from a single and dual-microphone setup on contact location estimation in Table II. We trained our model using data from either just one or two microphones, and evaluated the model's performance on the remaining data from the held-out microphones. The results show that training on a single-microphone resulted in accuracies from 94.7–99.4% (mean: 97.0%), and training on a dual-microphone setup resulted in accuracies from 98.3–99.2% (mean: 98.8%). The results show that incorporating multiple microphones can yield modest yet measurable improvements in system performance.

Object classification. We evaluate the planar sensor's ability to classify between the six objects. In each trial, the object is grasped at the center point of the surface, touching the membrane only, without any contact on the frame. Ten trials were recorded for each object. We performed 5-fold cross validation; the resulting accuracy was 97.7 $\pm$ 1.6% (Fig. 10).



(a) X-axis



(b) Y-axis

Fig. 9: **Spatial accuracy of planar tactile sensor.** Linear regression performed on the acoustic responses along the (a) x-axis and (b) y-axis of the planar sensor show RMSE values of 0.62 mm and 0.57 mm, respectively.

C. Benchmarks

Effect of different contact depths. We evaluate the effect of different contact depths on system performance for the planar sensor. To do this, we divide the planar sensor surface into a 3×3 grid with a spatial separation of 6 mm between adjacent points and collect a dataset of acoustic responses of five trials at three different depths at 1, 3, and 6 mm, for a total of 15 trials. After performing 5-fold cross-validation with an SVM model with linear kernel, the accuracy was 99.0 $\pm$ 1.5%. To assess how well the planar sensor could generalize to different contact depths, we trained the classifier on acoustic responses from two contact depths and evaluated it on responses from the third contact depth. The results are shown in Table III and have a system performance that ranges from 84.0 to 91.1%.

Effect of background noise. We evaluate the effect of background noise on the planar sensor classification performance. To do this, we repeat the contact estimation experiment by collecting a dataset of acoustic responses on the 3×3 grid of the planar sensor with a spacing of 6 mm. We collect 10 trials at each point while playing white noise at three different

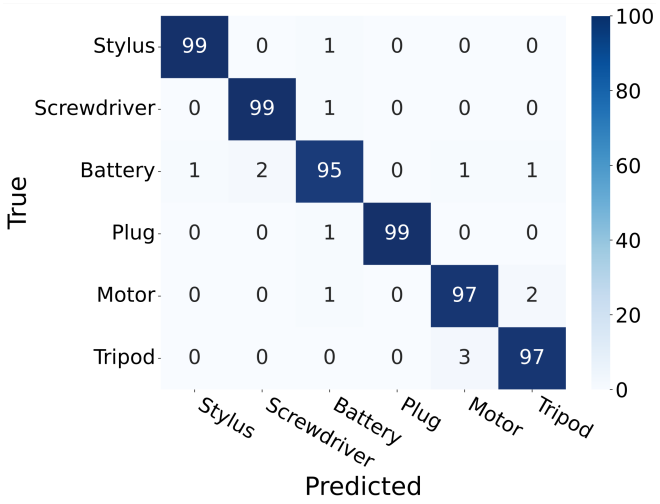


Fig. 10: **Confusion matrix for classifying 6 objects with planar tactile sensor.**

TABLE III: Planar sensor contact location estimation performance when tested on an unseen contact depth.

Train contact depth (mm)	Test contact depth (mm)	Accuracy (%)
1, 3	6	84.2
1, 6	3	91.1
3, 6	1	84.0

background sound levels: 70, 80, 86 dBA. When performing 5-fold cross-validation using an SVM within data collected at each of the sound level volumes, the accuracies were 100, 99.9, and 97.4% respectively. To evaluate the sensor's ability to generalize across different noise levels, we train the classifier on acoustic responses collected with two different background noise levels, and evaluate its performance collected at a different background noise level. Table IV shows that the classifier achieves a similar performance of 97–99.9% across the three combinations evaluated.

Effect of classifier. We examine the effect of different classifiers on contact location estimation for the linear and planar sensor using four additional models: Random Forest, Decision Tree, k-Nearest Neighbor, and Naïve Bayes. Table V shows that, for the linear sensor, the classifiers produce location estimation accuracies ranging from 84.7–93.9%, while the SVM with a linear kernel that we used for the main experiment produced an accuracy of 97.8%. For the planar sensor, the classifiers produced contact estimation accuracies ranging from 80.2–96.0%, while the SVM with a linear kernel produced an accuracy of 99.7%. These results show that the SVM produced the highest accuracy out of the classifiers evaluated.

Effect of transmission signal. We evaluate the planar sensor's contact location estimation performance on different transmission signals. We created two new signals; both signals were inaudible chirp sequences, consisting of a frequency sweep ranging from 18–20 kHz with duration 100 ms. The first signal was a linear frequency sweep, and the second signal was an exponential frequency sweep. In

TABLE IV: Planar sensor contact location estimation performance when tested on an unseen level of ambient noise.

Train sound levels (dBA)	Test sound levels (dBA)	Accuracy (%)
69.9, 80.0	85.7	97.0
80.0, 85.7	69.9	99.7
69.9, 85.7	80.0	99.9

TABLE V: Effect of classifier type on contact point location estimation accuracy.

Classifier type	Linear sensor accuracy (%)	Planar sensor accuracy (%)
Random Forest	94.0	96.0
Decision Tree	84.7	80.2
k-Nearest Neighbor	97.3	93.8
Naïve Bayes	89.3	84.2
SVM (linear kernel)	97.8	99.7

the experimental procedure, we transmitted these signals inside the sensor instead of the original pulse sequence and evaluated this new setup on the 3-by-3 grid described earlier. We performed 5-fold cross-validation. The accuracy of our system was $97 \pm 2.5\%$ and $99.2 \pm 0.8\%$, with the linear and exponential chirp, respectively.

V. Discussion & future work

Material sensing using acoustic vibrometry. Using the active acoustic sensing approach, our sensor detects objects based on deformations in its geometry, responding to differences in their structural features. As such, the sensor is not able to distinguish between geometrically identical objects made of different materials or internal structures. A future version of the sensor could leverage the principle of acoustic vibrometry, a technique used to sense product integrity [35]. In the context of our sensing system a surface transducer and microphone could be placed flush against the wall of the rubber tube or membrane which grasps the object, and send a wideband pulse or chirp to characterize the object's material. This procedure mimics the acoustic response when tapping against the object, but is non-invasive and can be used to characterize fragile and delicate objects without breaking it.

Increasing sensor sensitivity, spatial resolution, and durability Tactile sensors are characterized by their sensitivity, which indicates the minimum detectable pressure, and their spatial resolution, which defines the smallest area of depression required to register a change in state. Other low-cost elastomeric membranes made out of materials such as silicone have higher sensitivity and spatial resolution compared to our current design, and they have been used to detect surface characteristics like braille [36], using an internal camera and a complex assembly of hydraulic channels made of liquid-filled glass capillaries. Future work could focus on recording data on a higher-resolution grid for material sensing, comparing our sensor's payload capacity with other tactile sensors, enhancing membrane durability, and formally evaluating its lifespan over repeated use.

Few-shot calibration for contact point estimation. The deformation of our sensor produces complex, irregular shapes that are challenging to model using closed-form equations. These deformations alter the sensors internal geometry and introduce non-linear changes to its acoustic response. Although leveraging the entire acoustic response for contact point estimation along the planar sensor was more reliable than using manually selected features related to resonant frequencies, it required manual collection of calibration data for each contact point and can be challenging to scale.

An approach to reducing the amount of calibration data is to train a model that pairs simultaneously collected acoustic responses with camera data from within the sensor, which is able to visually encode contact location and depression shape from a smaller number of calibration points. Using a contrastive learning framework, the model can learn the shared latent space between the two modalities using positive and negative training pairs. At test time, the system can rely solely on the audio input and the learned latent space to reconstruct the likely visual deformation and perform few-shot contact point estimation on the planar sensor.

Microphone placement at nodes and antinodes. The current linear sensor design can be viewed as a resonant acoustic filter, with the microphone at the end recording the standing wave pattern at the antinode. Having multiple microphones along the length of the sensor, at multiple node and antinode points, can be used to provide complementary information that characterizes features of the standing wave and differentiates between subtle perturbations of the sensor. Specifically, microphones at the antinode capture high-pressure variations where constructive interference occurs at the resonant frequencies, while at the node there is low pressure variation due to destructive interference and nulls are observed at the resonant frequencies. The use of multiple microphones along soft robot end-effectors has been used for proprioception and shape reconstruction from the multiple acoustic signatures [23].

In conclusion, we present proof-of-concept, low-cost soft tactile sensor designs for contact point estimation and object classification using active acoustic sensing. The main contributions comprise the design of a linear tactile sensor to a robotic gripper as well as the development of a planar sensor featuring a novel multi-microphone recording setup. Our approach focuses on sensors that can be easily retrofitted onto existing robotic end-effectors and readily fabricated by researchers from various disciplines.

References

[1] *Robotiq 2f-85 gripper end-effector*, https://robotiq.com/products/adaptive-grippers#Two-Finger-Gripper, 2025.

[2] A. Alspach *et al.*, "Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation," in *RoboSoft*, IEEE, 2019.

[3] N. Kuppuswamy *et al.*, "Soft-bubble grippers for robust and perceptive manipulation," in *IROS*, IEEE, 2020.

[4] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, 2017.

[5] I. H. Taylor, S. Dong, and A. Rodriguez, "Gelslim 3.0: High-resolution measurement of shape, force and slip in a compact tactile-sensing finger," in *ICRA*, IEEE, 2022.

[6] B. Ward-Cherrier *et al.*, "The TacTip family: Soft optical tactile sensors with 3D-printed biomimetic morphologies," *Soft robotics*, 2018.

[7] A. Padmanabha *et al.*, "Omnitact: A multi-directional high-resolution touch sensor," in *ICRA*, IEEE, 2020.

[8] M. Lambeta *et al.*, "Digitizing touch with an artificial multimodal fingertip," *arXiv:2411.02479*, 2024.

[9] F. Zhu *et al.*, "Visual-tactile sensing for real-time liquid volume estimation in grasping," in *IROS*, IEEE, 2022.

[10] M. Y. Cao, S. Laws, and F. R. y Baena, "Six-axis force/torque sensors for robotics applications: A review," *IEEE Sensors*, 2021.

[11] V. Wall, G. Zöller, and O. Brock, "A method for sensorizing soft actuators and its application to the rbo hand 2," in *ICRA*, IEEE, 2017.

[12] N. Farrow and N. Correll, "A soft pneumatic actuator that can sense grasp and touch," in *IROS*, IEEE, 2015.

[13] H. Zhao *et al.*, "Optoelectronically innervated soft prosthetic hand via stretchable optical waveguides," *Science robotics*, 2016.

[14] R. Bhirangi *et al.*, "Reskin: Versatile, replaceable, lasting tactile skins," *arXiv:2111.00071*, 2021.

[15] Y. Luo *et al.*, "Learning human–environment interactions using conformal tactile textiles," *Nature Electronics*, 2021.

[16] J. Egli *et al.*, "Sensorized soft skin for dexterous robotic hands," in *ICRA*, IEEE, 2024.

[17] S. Zhang *et al.*, "A compact visuo-tactile robotic skin for micron-level tactile perception," *IEEE Sensors*, 2024.

[18] H. Li *et al.*, "See, hear, and feel: Smart sensory fusion for robotic manipulation," *arXiv:2212.03858*, 2022.

[19] K. Takaki *et al.*, "Acoustic length sensor for soft extensible pneumatic actuators with a frequency characteristics model," *IEEE RA-L*, 2019.

[20] G. Zöller, V. Wall, and O. Brock, "Active acoustic contact sensing for soft pneumatic actuators," in *ICRA '20*.

[21] V. Wall and O. Brock, "A virtual 2D tactile array for soft actuators using acoustic sensing," in *IROS '22*.

[22] K. Randika and K. Takemura, "Estimating the shape of soft pneumatic actuators using active vibroacoustic sensing," in *IROS*, IEEE, 2021.

[23] U. Yoo *et al.*, "POE: Acoustic soft robotic proprioception for omnidirectional end-effectors," *arXiv:2401.09382*, 2024.

[24] X. Fan *et al.*, "Aurasense: Robot collision avoidance by full surface proximity detection," in *IROS*, IEEE, 2021.

[25] S. Lu and H. Culbertson, "Active acoustic sensing for robot manipulation," in *IROS*, IEEE, 2023.

[26] S. Mikogai, B. Kazumi, and K. Takemura, "Contact point estimation along air tube based on acoustic sensing of pneumatic system noise," *IEEE RA-L*, 2020.

[27] M. Du *et al.*, "Play it by ear: Learning skills amidst occlusion through audio-visual imitation learning," *arXiv:2205.14850*, 2022.

[28] Z. Liu *et al.*, "ManiWAV: Learning robot manipulation from in-the-wild audio-visual data," in *8th Annual Conference on Robot Learning*, 2024.

[29] *1/4 in. I.D. x 3/8 in. O.D. x 10 ft. natural latex tubing*, https://www.homedepot.com/p/UDP-1-4-in-I-D-x-3-8-in-O-D-x-10-ft-Natural-Latex-Tubing-T64006001/304185180, 2025.

[30] *Sony MDREX15AP in-ear earbud headphones with mic, black*, https://www.amazon.com/Sony-MDREX15AP-B-Black/dp/B00JG2WRUO, 2025.

[31] *Analog microphone electret condenser*, https://www.digikey.com/en/products/detail/same-sky-formerly-cui-devices/CMC-6027-32L100/7398909, 2025.

[32] *Fielect 4pcs DIY magnetic speaker 0.5W 8 ohm 10mm diameter round shape replacement loudspeaker*, https://www.amazon.com/dp/B082762Y5W, 2025.

[33] *Cerrxian 3.5mm 4 pole stereo trrs audio video male to 4 screw terminal female headphone replacement solderless balum converter adapter (2-pack)*, https://www.amazon.com/Cerrxian-Terminal-Headphone-Converter-Adapter/dp/B06W2K9XMM.

[34] R. C. Snell and F. Milinazzo, "Formant location from lpc analysis data," *IEEE transactions on Speech and Audio Processing*, 1993.

[35] A. Gadre *et al.*, "MiLTOn: Sensing product integrity without opening the box using non-invasive acoustic vibrometry," in *IPSN*, IEEE, 2022.

[36] G. P. Jenkinson, A. T. Conn, and A. Tzemanaki, "Braille-tip: Structured small-footprint tactile sensor for high acuity dynamic tactile tasks," in *BioRob*, IEEE, 2024.