2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

POE: Acoustic Soft Robotic Proprioception for Omnidirectional End-effectors

Uksang Yoo [1,✉], Ziven Lopez[2] , Jeffrey Ichnowski[1*], and Jean Oh[1*]

***Abstract*— Shape estimation is crucial for precise control of soft robots. However, soft robot shape estimation and proprioception are challenging due to their complex deformation behaviors and infinite degrees of freedom. Their continuously deforming bodies complicate integrating rigid sensors and reliably estimating its shape. In this work, we present Proprioceptive Omnidirectional End-effector (POE), a tendon-driven soft robot with six embedded microphones. We first introduce novel applications of 3D reconstruction methods to acoustic signals from the microphones for soft robot shape proprioception. To improve the proprioception pipeline's training efficiency and model prediction consistency, we present POE-M. POE-M predicts key point positions from acoustic signal observations and uses an energy-minimization method to reconstruct a physically admissible high-resolution mesh of POE. We evaluate mesh reconstruction on simulated data and the POE-M pipeline with real-world experiments. Ablation studies suggest POE-M's guidance of the key points during the mesh reconstruction process provides robustness and stability to the pipeline. POE-M reduced the maximum Chamfer distance error by 23.1 % compared to the state-of-the-art end-to-end soft robot proprioception models and achieved 4.91 mm average Chamfer distance error during evaluation. Supplemental materials, experiment data, and visualizations are available at sites.google.com/view/acoustic-poe.**

I. Introduction

Soft robots have advantages in applications such as fruit harvesting [1, 2], food packaging [3], minimally invasive surgery [4, 5], contact-rich human-robot interactions [6], and human-wearable robot interactions [7, 8]. Soft robots' constituent deformable and elastic material make them well-suited for handling delicate objects and preventing harmful interactions [9]. Furthermore, recent works demonstrated that their compliance makes soft robots more robust in contact-rich control tasks [10, 11].

Soft robotic shape representation and estimation, or *proprioception*, is fundamental to many manipulation tasks but has challenges rooted in soft robots' inherent deformability and under-actuation [12, 13]. Observing the robot's state externally, which is a common alternative to proprioception, limits the robot's workspace to the external sensor's field of perception and is subject to occlusion [14, 15]. Prior research often fit low degree-of-freedom shape primitives, such as constant-curvature curves to represent the centerline of the robot [16, 17]. Such approaches, however, filter out complex deformation behaviors by reducing the independent degrees of freedom [13].

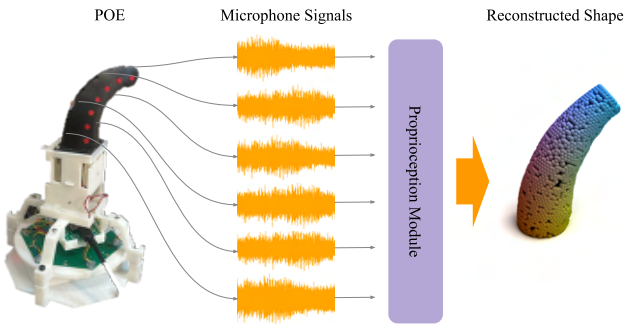

Fig. 1. Overview of the proposed pipeline for Proprioceptive Omnidirectional End-effector (POE) acoustic soft robotic proprioception. We obtain acoustic signals from our novel tendon-driven soft robot POE with six embedded microphones (left). We then feed the extracted acoustic features into our proprioception modules that can reconstruct a high degree of freedom shape of POE (right).

Internally embedded cameras can capture complex deformation behavior if the soft robot's internal lighting and spatial conditions allow for vision-based proprioception [13, 18]. In these applications, vision provides rich observations of surface deformations [13]. However, vision-based approaches require specially designed robots with a sufficiently large cavity to limit self-occlusion. Furthermore, these methods generally use vision models that require large training datasets of image-shape pairs.

Acoustic sensing is an attractive alternative for soft robot proprioception because microphones can be small, scaled to an arbitrary array size, and be installed with few steps. Thus, we propose acoustically Proprioceptive Omnidirectional End-effector (POE). POE has an array of embedded microphones to observe changes in the acoustic propagation properties of the soft robot with deformation. We test with baseline reconstruction pipelines based on K-nearest neighbors (KNN) and an encoder-decoder network inspired by DeepSoRoNet [12, 13] that signals from embedded microphone arrays can reconstruct the robot shape reliably in diverse loading conditions. Furthermore, we introduce POE-M, a pipeline that predicts key points that guide a mesh reconstruction module toward full POE shape estimation. In real-world experiments, POE-M reconstructs the state of the soft robot with 23.1 % lower maximum Chamfer distance error compared to learning to directly reconstruct a point cloud.

This paper makes the following contributions:

✉ Corresponding author.
* Equal contribution
[1]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA {uyoo, jichnows, hyaejino}@andrew.cmu.edu
[2]Northeastern University, Boston, MA, 02115, USA {lopez.z}@northeastern.edu

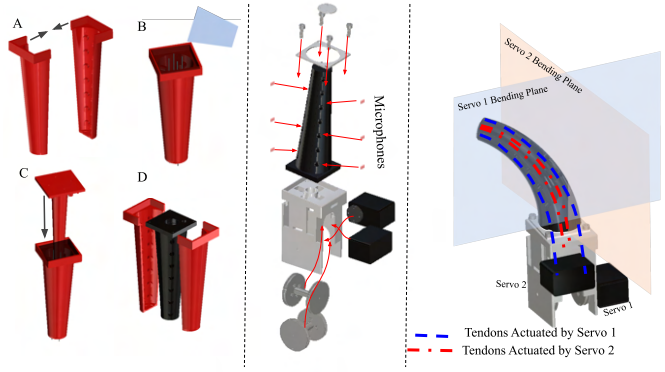

Fig. 2. POE is a tendon-driven soft robot fabricated with a single molding step. **Left column:** The three-part mold is assembled with four metal rods (B) before the uncured silicone is poured in to create channels for tendons routing. The last part of the mold is then inserted to make the central conical cavity (C). After curing, the mold is disassembled to get the POE finger. **Middle column:** the finger secured into the rest of POE assembly with two servo motors for tendon actuation. **Right column:** each servo controls POE movement in its perpendicular plane of bending, enabling POE to bend toward any direction.

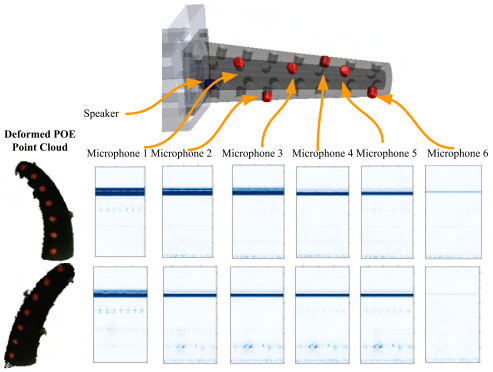

Fig. 3. Spectogram visualizations from the six embedded microphones in two different shape configurations. The signal magnitudes change from one shape to another as we observe from the two rows of spectrograms. Time-varying features are ignored in the presented pipelines by averaging over the recording.

1) Design and evaluation of a tendon-driven soft robot with embedded microphones and an active sound source,
2) POE-KNN and POE-DeepSoRo baseline pipelines that demonstrate the efficacy of using acoustic signals for soft robot proprioception,
3) POE-M pipeline which approximates key points from acoustic signals and reconstructs a watertight mesh of the deformed robot with a physics-inspired energy-minimization method,
4) Acoustic-shape dataset for POE with an evaluation of the proposed pipelines,
5) Demonstration of using the POE-M pipeline for shape feedback control.

II. Related Work

A. Soft Robotic Manipulators

Design and development of soft-robotic manipulators are active fields of research with growing interest [19, 20]. The mode of actuation for soft robots varies widely from dielectric artificial muscle to pneumatic. Pneumatically actuated and tendon-driven soft robots generally only require readily available materials [16, 21]. The relative fabrication simplicity of pneumatic and tendon-driven soft robots have made them popular choices [22].

B. Soft Robotic Proprioception

To address challenges in soft robot state estimation and proprioception, researchers have proposed various methods to directly measure deformations [23]. A popular approach embeds flexible and elastic strain sensors along the robot length [24–26]. However, implementing these direct strain sensing methods can be difficult and expensive since they can require specialized sensors, hardware, and domain expertise to fabricate [25, 27]. Furthermore, the complexity of soft robot design grows substantially with an increasing number of these sensing elements, requiring robot-specific sensor placement optimization [28]. These works have also focused on simplifying to the limited degree of freedom representations [24]. To capture soft robots' complex deformation behaviors, we propose POE pipelines that represent shapes with dense point clouds.

Indirect methods of soft robotic proprioceptive sensing do not measure local strain and deformation directly. A popular approach embeds a camera inside the soft robot to observe internal surface deformation and infer the external shape [12, 13, 29, 30]. Because vision provides rich observations of the robot's state, some previous works were compatible with high degree-of-freedom soft-robot shape representations [12]. These methods generally rely on an encoder-decoder neural network architecture to learn a latent representation of the robot that decodes to the full robot shape represented by point clouds. In the process, physics-based constraints must be learned implicitly from examples, requiring a large training dataset of image-shape pairs to consistently yield physically admissible soft-robot shape predictions [13]. In this paper, we propose addressing these challenges with stronger supervision by predicting key point movements on the mesh and framing shape decoding as a mesh energy optimization problem [31].

C. Acoustic Sensing for Soft Robots

Arrays of microphones can be embedded into soft bodies to provide indirect contact sensing in soft bodies [32]. For soft robots, acoustic sensing is advantageous because low-cost miniature microphones are readily available [33]. Furthermore, by using the soft robot's deformable body as the medium for the propagation of acoustic vibrations from an active sound source, microphones can detect changes to the material state such as strain [34]. Previous works showed that a microphone embedded in soft robots can be used for various applications ranging from contacting object material property classification, contact position estimation, temperature regression, and braille letter classification [33–35]. The previous works generally relied on a single embedded microphone and framed the tasks as a coarse classification or low degree-of-freedom regression problem [34]. To the

best of our knowledge, this work presents the first multi-microphone high-fidelity shape reconstruction of the robot using acoustic signals.

III. Methods

We outline methods for the design of POE and processing the acoustic signals. Then we discuss the formulation of the POE-M pipeline.

A. Design and Fabrication of POE

For potential applications such as dexterous manipulation, POE can be actuated toward any direction radially. We also developed POE to be low-cost and mechanically simple for reproducibility. We chose to develop POE with tendon-actuation for simplicity. POE is actuated by four tendons anchored at the fingertip and tensioned by two servos (Dynamixel, XC330-M288-T). POE's design allows each servo motor to control the robot pose in a bending plane (Fig. 2). The combination of the two servos can actuate POE to move its fingertip to any point in its semi-hemisphere workspace without contact.

POE's soft finger is molded with a 3-part mold with silicone rubber (Smooth-On, Inc. Dragon Skin^{TM} 20) to allow it to be elastic and deformable. The mold has four narrow cavities where the nylon tendons are inserted after curing. POE has a central conical cavity that allows it to bend without buckling and serves as a channel for the embedded microphone cables. POE is 110 mm long, making it similar in dimension to an average adult human's middle finger [36]. Six miniature microphones (ReSpeaker, Circular Array) are inserted into the preset cavities and secured with silicone adhesive. The base of POE's molded finger has an embedded miniature speaker (Knowles, RAB) to act as an active sound source. The acoustic signals that POE microphone arrays observe when POE is deformed into different shapes are shown in Fig. 3 with their corresponding microphone locations.

B. Audio Signal Processing

We processed the audio from the microphones to be used as input features into the rest of the reconstruction pipelines and to remove consistent background noise. At each POE pipeline's initialization, we record a 1.0-second background audio. We average the audio clip's spectrogram with Tukey window of shape parameter 0.25 over time to get the initial acoustic feature vector. In subsequent instances, we collected 0.25 seconds of acoustic signals from each of the six microphones. We then extract spectrograms from each of the acoustic signal clips and average over time. Then we subtract the initial spectrogram feature vector from each microphone's averaged spectrograms and the acoustic features from each of the microphones are concatenated.

C. POE-M

We propose POE-Multilayer perceptron (POE-M) reconstruction method. The key insight of POE-M is that by first estimating explicit key point movements from the sensor signals, we can provide strong supervision when we train the acoustic signal decoder and constrain soft robot deformation behaviors to physically admissible transformations.

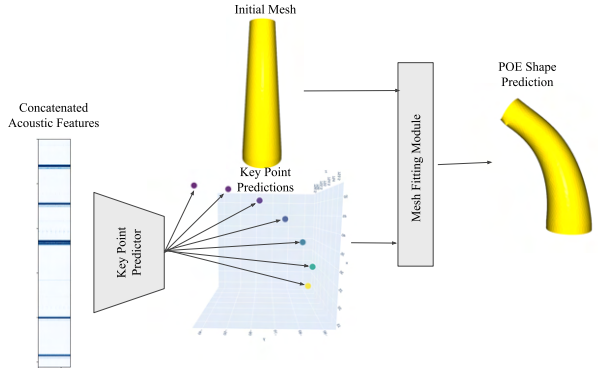

Fig. 4. Proposed POE-M Pipeline. First, concatenated acoustic feature vector from six microphones embedded in POE are used to predict new positions of the key points on POE's surface. POE-M uses the known correspondences between the key points and the vertices of POE surface mesh to iteratively fit the mesh to the predicted key points in a physically admissible manner.

1) Key Point Model: Key points provide physically grounded reduced state representation for high degree-of-freedom systems such as fabric [37] and ropes [38, 39]. For POE, using key points is also advantageous because they can provide the POE-M pipeline with stronger supervision during training in contrast to previous works [13] that relied on Chamfer distance loss which frequently leads to falling into local optima [40].

Another advantage to approximating key points as an explicit intermediate representation of POE state is that POE-M can exploit the key points' physically grounded correspondence to the vertices on POE's surface mesh. This allows POE-M to utilize the predicted sparse key point movements in reconstructing POE's deformed mesh. We chose to track seven key points evenly distributed along the midline of POE's outer surface to capture the POE deformation.

In our implementation of the POE-M key point predictor model, the input acoustic feature vector maps to the seven 3D key points with a two-layer MLP. The input acoustic feature from the microphone array has the dimension 512 and the output is vectorized 3D position changes of the key points.

2) Mesh Reconstruction: As-Rigid-As-Possible (ARAP) is a method of mesh manipulation popular in animation and graphics applications. It uses chosen *handle points* to deform meshes in a physically realistic and grounded way. Some ARAP formulations can also be applied to regularize visual reconstructions and introduce local rigidity to surfaces [42]. We treat the key points' corresponding vertices on the initial POE mesh as the handle points and use the predicted key point positions to infer handle point movements. We first define the source surface mesh S and the deformed mesh S', where each mesh is a set of vertices V, V' and of edges E, E'. We define the ARAP energy as [43]:

$$E_{\text{ARAP}}(S, S') = \sum_{k=1}^{|E|} \min_{R \in SO(3)} \sum_{e_{i,j} \in E} w_{i,j} \| e'_{i,j} - R e_{i,j} \|.$$

We then find the deformed mesh S' that minimizes E_{ARAP}

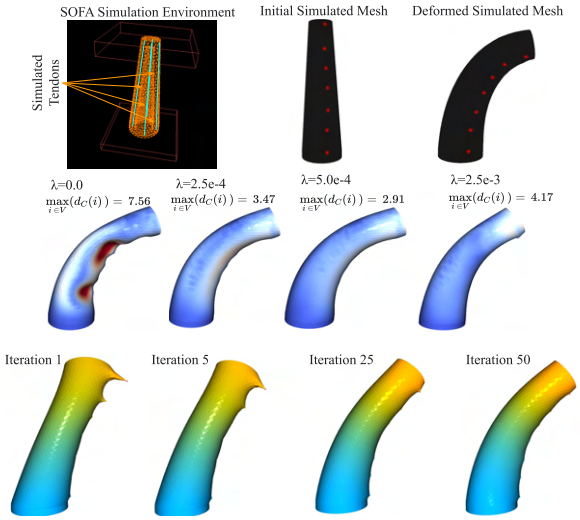

Fig. 5. Evaluation of the As-Rigid-As-Possible (ARAP). **Top row:** SOFA FEM simulation environment [41] and the generated meshes to be used for ARAP evaluation. **Middle row:** sensitivity study with varying λ parameter. When $\lambda = 0$ which corresponds to no neighboring edge rotation regularization, we observe undesirable surface artifacts. All nonzero λ parameters removed the artifact effectively where $\lambda = 5.0e{-}4$ yielded the lowest mesh reconstruction error. **Bottom row:** mesh updates over iterations. After each iteration, the mesh vertices that are not constrained are optimized to reduce the overall ARAP energy. We note that at around 30-50 iterations, the mesh converges.

with an iterative local-global optimizer from Levi, et al. [44]. It is a known issue that minimizing E_{ARAP} with sparse handle points on surface meshes that are moved significantly can result in undesirable surface artifacts.

A possible solution to this problem is by minimizing the E_{ARAP} over a tetrahedral mesh instead of a surface triangular mesh of POE, which implicitly applies soft volumetric constraints that prevent such artifacts from forming. However, because tetrahedral meshes by construction include numerous internal vertices and edges that must be operated over to minimize E_{ARAP}, the process becomes more computationally expensive which is especially undesirable in the context of doing real-time soft robotic proprioception.

Instead, prior works showed that modification of ARAP to include a penalty on the rotations of the neighboring edges produces mesh manipulation that seems physically admissible [44]. The new energy to minimize is

$$E_{\text{smoothed}}(S, S') = \min_{R_1,\ldots,R_m} \sum_{k=1}^{m} \Big(\sum_{i,j \in e_k} c_{ijk} \|e_{ij} - R_k e_{ij}\|^2 + \lambda \hat{A} \sum_{e_l \in N(e_k)} w_{kl} \|R_k - R_l\|^2 \Big).$$

Note that minimizing E_{smoothed} with $\lambda = 0$ results in the minimizing of E_{ARAP}. In the POE-M pipeline, vertices corresponding to the key points $p_{1,\ldots,|p_k|}$ are constrained to the new positions based on the predicted key-point positions, and the rest of the mesh vertex positions are moved to minimize E_{smoothed}. As outlined in Fig. 4, the key point predictor and the mesh fitting module which uses the smoothed ARAP formulation are connected to enable full POE shape prediction from the acoustic signal feature vector.

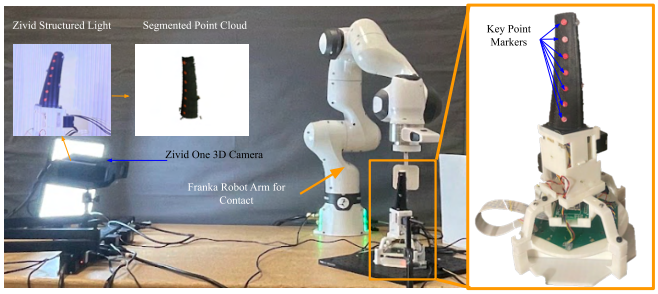

Fig. 6. Experimental setup for collecting training and evaluation data. We use a structured light 3D camera to capture high-fidelity point clouds of POE as it deforms. **Right:** POE has embedded key point markers that are then used to train POE-M to extract key point displacements from acoustic signals.

TABLE I
PROPRIOCEPTIVE PERFORMANCE EVALUATION

Model	Model Input	Performance Metrics [mm] Avg. CD ↓	Max CD ↓
Position-KNN	Servo Positions	5.67	20.21
POE-KNN	Audio [All]	5.93	17.02
POE-DeepSoRo	Audio [All]	**4.89**	15.33
POE-M	Audio [-mic1]	5.46	13.11
POE-M	Audio [-mic6]	5.05	**11.79**
POE-M	Audio [All]	4.91	**11.98**

IV. EVALUATION

A. Baselines

As a baseline for acoustic signal, we included a nearest neighbor model that takes in only the two servo motors' encoder positions and matches them to the closest seen example and its corresponding observed POE shape. The rest of the POE family of methods use acoustic signals from the microphone arrays; the results with each suggest that acoustic signal is a viable occlusion-free modality for soft robot proprioception. First, we present *POE-KNN* which matches POE acoustic signal to its nearest neighbor in previously seen examples and use the corresponding point-cloud observation as its prediction. We then present *POE-DeepSoRo* in which we adapt a DeepSoRo decoder-encoder network architecture from vision to acoustics [12, 13].

B. Experimental Setup

To evaluate the 3D reconstruction results of the POE proprioceptive pipelines and to collect diverse training and evaluation datasets in the real world, we set up the experimental environment as shown in Fig. 6. The structured light 3D camera (Zivid, One Plus) provides us with low-noise point clouds with 25 μm accuracy which serve as a sort of ground truth to evaluate our methods. The 7 degrees-of-freedom robot arm is also installed such that it can create various contact conditions for POE to deform from in addition to POE's tendon-driven range of motion. Such loading conditions allow us to evaluate the various reconstruction pipelines within the expanded configuration space. We collected 5200 point-cloud audio pairs with this

setup and then split 80:20 into training and testing sets respectively. Each point cloud is segmented with HSV color thresholding. For Position-KNN, we also collected servo encoder positions for the two servo motors. For POE-M models, we also segmented the red markers indicating the key point positions on POE. Diffused lighting was added for consistent background illumination and segmentation results. Note that the 3D camera and controlled environment for POE are only required for training and evaluation data collection, and are not requirements for the operation.

C. Simulation Study

Despite ARAP's wide usage in animation and graphics, its application in modeling real-world mechanics accurately is rarely explored. Furthermore, to the best knowledge of the authors, it has never been applied to modeling deformations of actuated soft robots. The later sections present the real-world performance of POE-M and baselines. However, obtaining an occlusion-free point cloud of any deforming objects is notably difficult, and disambiguating sensor noise from the method-inherent errors is difficult. To verify that the proposed POE-M's mesh fitting module performs satisfactorily, we test that the module produces results that are aligned with previously tested and verified Finite Element Method (FEM)-based physics simulators. Specifically, we simulated POE mechanics with SOFA framework [41] which has been repeatedly demonstrated in literature to match real-world elastic and soft mechanics [13].

With the FEM simulator, we generated a deformed mesh of POE under actuation. In the process, we also track how the key points corresponding to the key points on real POE moved. Based on the displacements of the key points, we apply our mesh fitting module on the initial mesh of POE to match it to the deformed mesh. We also study the sensitivity of the λ smoothening parameter from $E_{smoothed}$. When $\lambda = 0$, $E_{smoothed}$ reduces to minimization of original ARAP energy (E_{ARAP}), and we encounter undesirable surface artifacts with a maximum Chamfer distance of 7.56 mm. With the introduction of a nonzero λ, there is a significant improvement in the Chamfer distance with the best-tested value $\lambda = 5 \times 10^{-4}$, yielding a maximum Chamfer distance of 2.91 mm.We can observe generally low sensitivity of the mesh-fitting module performance with respect to λ values, where each nonzero λ value tested performed well. For subsequent experiments, we set $\lambda = 5 \times 10^{-4}$.

D. Metrics

In this work, we propose and report two metrics that must be considered together: average unidirectional Chamfer distance and maximum unidirectional Chamfer distance. The proposed POE-M pipeline generates watertight meshes with dense vertices evenly distributed on POE's finger. However, our real-world experimental setup can only produce a partial point cloud of POE's surface. The unidirectional Chamfer distance is defined as

$$d_{\mathrm{UCD}}(S,T) = \frac{1}{|S|} \sum_{p_i \in S} \min_{p_j \in T} \|p_i - p_j\|_2. \qquad (1)$$

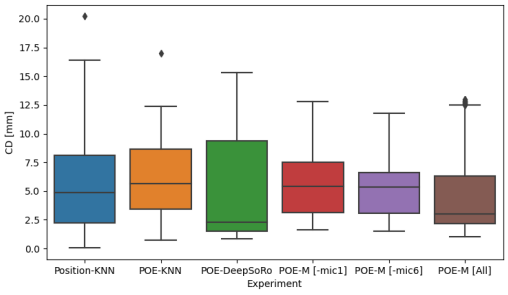

Fig. 7. The distribution of unidirectional Chamfer distance across the 1020 evaluation data points. The plot displays outliers marked by diamond markers and the quartiles from the evaluation dataset.

The intuition is that d_{UCD} measures how well the observed partial point cloud S fits the target complete predicted point cloud T. We report both the d_{UCD} averaged across different sampled shapes, and the maximum to capture both the overall and worst performance.

E. Results

Table I and Fig. 7 outlines the performances of both baseline and proposed shape reconstruction methods for POE. Position-KNN baseline method has a comparatively low mean (Table I) and median (centerline of Fig. 7). However, the quartile distribution and the maximum recorded Chamfer distance indicate that the baseline method is also prone to fail with commonly large errors (>15.0 mm) and a maximum error of 20.21 mm. Position-KNN result highlights the primary challenge in proprioception for soft robots, that is, the compliance and underactuation of the soft robots mean that the tendons can not fully constrain the pose. Additionally, tendon-driven robots suffer from various effects of hysteresis and pose-dependent internal tendon friction which alters the soft robot shape even with the same tendon positions [45]. We can observe these failures with POE-KNN and Position-KNN in Fig. 8 where we see large prediction deviations, especially with contact.

As a direct comparison against Position-KNN–which only uses servo encoder positions for soft robot state estimation, the performance of POE-KNN illustrates similarly low average Chamfer distance (< 6.0 mm) but its range of errors was much tighter with significantly lower maximum Chamfer distance at 15.8 % below that of Position-KNN and a sharply lower outlier-removed maximum Chamfer distance as seen in Fig. 7. Qualitatively, POE-KNN performs well in densely sampled regions as can be seen in Fig. 8.

POE-DeepSoRo demonstrates an exceptionally good performance with a low average Chamfer distance (< 5.0 mm). However, its errors significantly increased in some bending regions, where its Q3 quartile boundary is higher than the baseline methods. This result highlights one of the disadvantages of end-to-end learning for soft robot proprioception in that it learns point-to-point relationships entirely from training examples with no mechanics-based grounding. The result shows that this potentially allows the model to make physically impossible transformations on the point cloud.

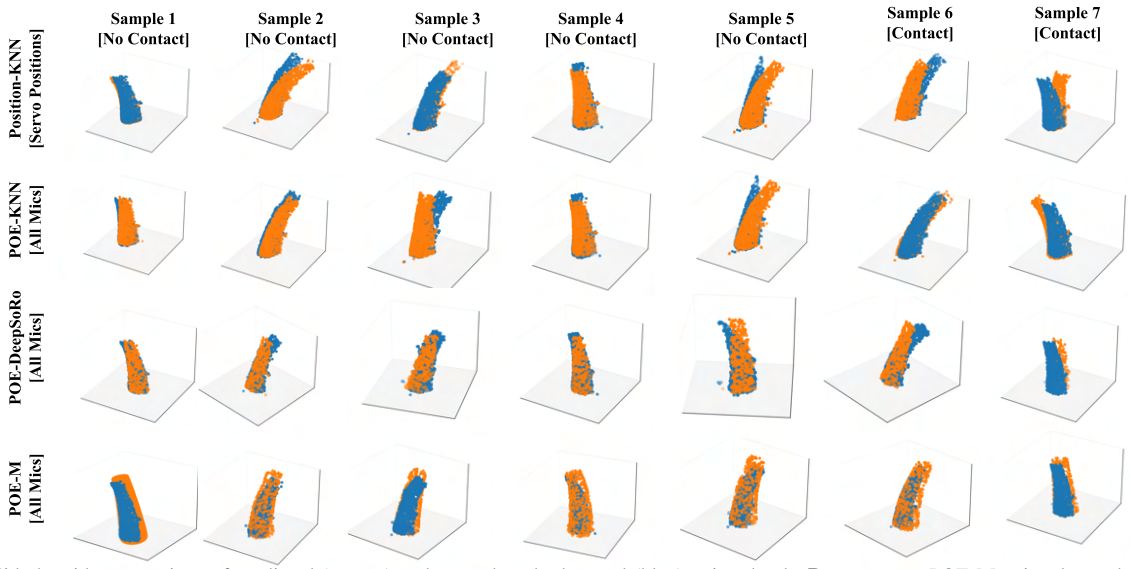

Fig. 8. Side-by-side comparison of predicted (orange) and ground-truth observed (blue) point clouds. **Bottom row**: POE-M uniquely produces a complete point cloud of the deformed POE shape. POE-M also produces generally more stable shape estimates in contrast to POE-DeepSoRo which produces shape estimates with arbitrarily deformed morphology. Both POE-KNN and Position-KNN predictions occasionally fail with large deviations from the ground truth, notable when POE is in contact (samples 6 and 7).

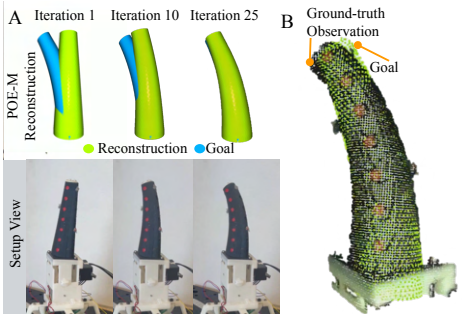

Fig. 9. Demonstration of using POE-M acoustic reconstruction pipeline for shape feedback control. We use a proportional controller to actuate the robot to the desired goal shape until it converges using only the reconstructed shape for feedback (A). The goal shape closely matches the ground-truth observed shape (B).

POE-M outperforms the baselines as it preserves the mesh connectivity and returns the full shape estimation of the robot (see Fig. 8). This is a crucial advantage of POE-M in that the method can predict even the unobserved POE surface. Furthermore, by framing the soft-robot mesh fitting process as an optimization problem over the edges of the surface mesh, we are instilling stability to the soft surface reconstruction problem [42]. We can note that the results look remarkably consistent with POE's mechanics. These built-in advantages in POE-M are visible in Table I and Fig. 7 where POE-M has consistently lower variance in error compared to all other methods.

When POE-M is retrained with a single microphone signal dropped out, the error remains consistently low. Interestingly, we note from Table I that dropping microphone 1's channel seemed to have a much more significant negative impact than dropping microphone 6's channel did. This is potentially explained by the fact that microphone 6 is much further away from the speaker, inducing much weaker signals and perceivable changes as the robot deforms.

Finally, we present a case study demonstration of using the POE-M pipeline for a proportional feedback controller. We define the goal state with a deformed POE mesh. Using the correspondences between the goal mesh and the reconstructed mesh from POE-M pipeline, we find the average of all of the vertices' residuals denoted by vector $\bar{r}$. We then define the rotation action in radians for the servo motor as $u_i = k\bar{r} \cdot e_i$ where $k = 0.005$ is a scalar control gain used in this experiment and e_i is the control direction vector on the bending plane in Fig. 2 of servo motor i. After 30 iterations, we observed that the ground-truth observation closely matched the goal shape with $d_{UCD} = 5.93$ as shown in Fig. 9B.

V. Conclusion

We present POE, the first soft robotic system with six embedded microphones that enable acoustic-based high-fidelity proprioception. We also present and evaluate a family of methods that utilize these acoustic signals for shape reconstruction and proprioception. Of these, we discuss the benefits of representing states with physically grounded key points and reconstructing the mesh around the key point movements in a physically admissible manner. We quantitatively and qualitatively verified that introducing these soft mechanics-based constraints enables more stable and accurate proprioceptive shape reconstruction results. We plan to address the current pipeline's limitation of cross-talk among multiple active sound sources by applying a range of sound source disambiguation methods in literature [46].

Acknowledgements

This work is supported by NSF Graduate Research Fellowship under Grant No. DGE2140739 and by the Technology Innovation Program (20018112, Development of autonomous manipulation and gripping technology using imitation learning based on visual tactile sensing) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).

References

[1] A. L. Gunderman, J. A. Collins, A. L. Myers, R. T. Threlfall, and Y. Chen, "Tendon-driven soft robotic gripper for blackberry harvesting," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2652–2659, 2022. 1

[2] A. L. Gunderman, J. A. Collins, A. L. Myers, R. T. Threlfall, and Y. Chen, "Tendon-driven soft robotic gripper for blackberry harvesting," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2652–2659, 2022. 1

[3] J. H. Low, P. M. Khin, Q. Q. Han, H. Yao, Y. S. Teoh, Y. Zeng, S. Li, J. Liu, Z. Liu, P. V. y Alvarado, *et al.*, "Sensorized reconfigurable soft robotic gripper system for automated food handling," *IEEE/ASME Transactions On Mechatronics*, vol. 27, no. 5, pp. 3232–3243, 2021. 1

[4] M. Runciman, A. Darzi, and G. P. Mylonas, "Soft robotics in minimally invasive surgery," *Soft robotics*, vol. 6, no. 4, pp. 423–443, 2019. 1

[5] K.-W. Kwok, H. Wurdemann, A. Arezzo, A. Menciassi, and K. Althoefer, "Soft robot-assisted minimally invasive surgery and interventions: Advances and outlook," *Proceedings of the IEEE*, vol. 110, no. 7, pp. 871–892, 2022. 1

[6] U. Yoo, N. Dennler, M. Mataric, S. Nikolaidis, J. Oh, and J. Ichnowski, "MOE-Hair: Toward Soft and Compliant Contact-rich Hair Manipulation and Care," in *Proceedings of the Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '24, ACM, 2024. 1

[7] C. Walsh, "Human-in-the-loop development of soft wearable robots," *Nature Reviews Materials*, vol. 3, no. 6, pp. 78–80, 2018. 1

[8] M. Zhu, S. Biswas, S. I. Dinulescu, N. Kastor, E. W. Hawkes, and Y. Visell, "Soft, wearable robotics and haptics: Technologies, trends, and emerging applications," *Proceedings of the IEEE*, vol. 110, no. 2, pp. 246–272, 2022. 1

[9] T. Watanabe, K. Yamazaki, and Y. Yokokohji, "Survey of robotic manipulation studies intending practical applications in real environments-object recognition, soft robot hand, and challenge program and benchmarking," *Advanced Robotics*, vol. 31, no. 19-20, pp. 1114–1132, 2017. 1

[10] A. Bhatt, A. Sieler, S. Puhlmann, and O. Brock, *Surprisingly robust in-hand manipulation: An empirical study*. 2022. 1

[11] A. Sieler and O. Brock, "Dexterous soft hands linearize feedback-control for in-hand manipulation," 2023. 1

[12] R. Wang, S. Wang, S. Du, E. Xiao, W. Yuan, and C. Feng, "Real-time soft body 3d proprioception via deep vision-based sensing," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3382–3389, 2020. 1, 2, 4

[13] U. Yoo, H. Zhao, A. Altamirano, W. Yuan, and C. Feng, "Toward zero-shot sim-to-real transfer learning for pneumatic soft robot 3d proprioceptive sensing," 2023. 1, 2, 3, 4, 5

[14] D. Bruder, X. Fu, R. B. Gillespie, C. D. Remy, and R. Vasudevan, "Koopman-based control of a soft continuum manipulator under variable loading conditions," *IEEE robotics and automation letters*, vol. 6, no. 4, pp. 6852–6859, 2021. 1

[15] D. A. Haggerty, M. J. Banks, E. Kamenar, A. B. Cao, P. C. Curtis, I. Mezić, and E. W. Hawkes, "Control of soft robots with inertial dynamics," *Science robotics*, vol. 8, no. 81, p. eadd6864, 2023. 1

[16] U. Yoo, Y. Liu, A. D. Deshpande, and F. Alamabeigi, "Analytical design of a pneumatic elastomer robot with deterministically adjusted stiffness," *IEEE robotics and automation letters*, vol. 6, no. 4, pp. 7773–7780, 2021. 1, 2

[17] R. Szász, M. Allenspach, M. Han, M. Tognon, and R. K. Katzschmann, "Modeling and control of an omnidirectional micro aerial vehicle equipped with a soft robotic arm," in *2022 IEEE 5th International Conference on Soft Robotics (RoboSoft)*, pp. 01–08, IEEE, 2022. 1

[18] A. Zhang, R. L. Truby, L. Chin, S. Li, and D. Rus, "Vision-based sensing for electrically-driven soft actuators," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11509–11516, 2022. 1

[19] D. Rus and M. T. Tolley, "Design, fabrication and control of soft robots," *Nature*, vol. 521, no. 7553, pp. 467–475, 2015. 2

[20] E. W. Hawkes, C. Majidi, and M. T. Tolley, "Hard questions for soft robotics," *Science robotics*, vol. 6, no. 53, p. eabg6049, 2021. 2

[21] J. Fras, J. Glówka, and K. Althoefer, "Instant soft robot: A simple recipe for quick and easy manufacturing," in *2020 3rd IEEE International Conference on Soft Robotics (RoboSoft)*, pp. 482–488, IEEE, 2020. 2

[22] O. Yasa, Y. Toshimitsu, M. Y. Michelis, L. S. Jones, M. Filippi, T. Buchner, and R. K. Katzschmann, "An overview of soft robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 6, pp. 1–29, 2023. 2

[23] H. Wang, M. Totaro, and L. Beccai, "Toward perceptive soft robots: Progress and challenges," *Advanced Science*, vol. 5, no. 9, p. 1800541, 2018. 2

[24] V. Wall, G. Zöller, and O. Brock, "A method for sensorizing soft actuators and its application to the rbo hand 2," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4965–4970, IEEE, 2017. 2

[25] J. Tapia, E. Knoop, M. Mutnỳ, M. A. Otaduy, and M. Bächer, "Makesense: Automated sensor design for proprioceptive soft robots," *Soft robotics*, vol. 7, no. 3, pp. 332–345, 2020. 2

[26] J. So, U. Kim, Y. B. Kim, D.-Y. Seok, S. Y. Yang, K. Kim, J. H. Park, S. T. Hwang, Y. J. Gong, and

H. R. Choi, "Shape estimation of soft manipulator using stretchable sensor," *Cyborg and Bionic Systems*, 2021. 2

[27] H. Zhao, K. O'brien, S. Li, and R. F. Shepherd, "Optoelectronically innervated soft prosthetic hand via stretchable optical waveguides," *Science robotics*, vol. 1, no. 1, p. eaai7529, 2016. 2

[28] V. Wall and O. Brock, "Multi-task sensorization of soft actuators using prior knowledge," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 9416–9421, IEEE, 2019. 2

[29] M. Hofer, C. Sferrazza, and R. D'Andrea, "A vision-based sensing approach for a spherical soft robotic arm," *Frontiers in Robotics and AI*, vol. 8, p. 630935, 2021. 2

[30] Y. She, S. Q. Liu, P. Yu, and E. Adelson, "Exoskeleton-covered soft finger with vision-based proprioception and tactile sensing," pp. 10075–10081, 2020. 2

[31] A. Sanchez-Gonzalez, N. Heess, J. T. Springenberg, J. Merel, M. Riedmiller, R. Hadsell, and P. Battaglia, *Graph networks as learnable physics engines for inference and control*. 2018. 2

[32] K. Park, H. Yuk, M. Yang, J. Cho, H. Lee, and J. Kim, "A biomimetic elastomeric robot skin using electrical impedance and acoustic tomography for tactile sensing," *Science Robotics*, vol. 7, no. 67, p. eabm7187, 2022. 2

[33] G. Zöller, V. Wall, and O. Brock, "Active acoustic contact sensing for soft pneumatic actuators," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7966–7972, IEEE, 2020. 2

[34] V. Wall and O. Brock, "A virtual 2d tactile array for soft actuators using acoustic sensing," 2022. 2

[35] V. Wall, G. Zöller, and O. Brock, "Passive and active acoustic sensing for soft pneumatic actuators," *The International Journal of Robotics Research*, vol. 42, no. 3, pp. 108–122, 2023. 2

[36] M. Peters, K. Mackenzie, and P. Bryden, "Finger length and distal finger extent patterns in humans," *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, vol. 117, no. 3, pp. 209–217, 2002. 3

[37] C. He, L. Meng, J. Wang, and M. Q.-H. Meng, "Fabricfolding: Learning efficient fabric folding without expert demonstrations," *arXiv preprint arXiv:2303.06587*, 2023. 3

[38] X. Lin, Y. Wang, J. Olkin, and D. Held, "Softgym: Benchmarking deep reinforcement learning for deformable object manipulation," in *Conference on Robot Learning*, pp. 432–448, PMLR, 2021. 3

[39] J. Grannen, P. Sundaresan, B. Thananjeyan, J. Ichnowski, A. Balakrishna, V. Viswanath, M. Laskey, J. Gonzalez, and K. Goldberg, "Untangling dense knots by learning task-relevant keypoints," in *Conference on Robot Learning*, pp. 782–800, PMLR, 2021. 3

[40] Y. Li, A. Zeng, and S. Song, "Rearrangement planning for general part assembly," in *7th Annual Conference on Robot Learning*, 2023. 3

[41] P. Schegg, E. Ménager, E. Khairallah, D. Marchal, J. Dequidt, P. Preux, and C. Duriez, "Sofagym: An open platform for reinforcement learning based on soft robot simulations," *Soft Robotics*, vol. 10, no. 2, pp. 410–430, 2023. 4, 5

[42] Y. Li, F. Richter, J. Lu, E. K. Funk, R. K. Orosco, J. Zhu, and M. C. Yip, "Super: A surgical perception framework for endoscopic tissue manipulation with surgical robotics," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2294–2301, 2020. 3, 6

[43] O. Sorkine and M. Alexa, "As-rigid-as-possible surface modeling," in *Symposium on Geometry processing*, vol. 4, pp. 109–116, Citeseer, 2007. 3

[44] Z. Levi and C. Gotsman, "Smooth rotation enhanced as-rigid-as-possible mesh animation," *IEEE transactions on visualization and computer graphics*, vol. 21, no. 2, pp. 264–277, 2014. 4

[45] Y. Liu, U. Yoo, S. Ha, S. F. Atashzar, and F. Alambeigi, "Influence of antagonistic tensions on distributed friction forces of multisegment tendon-driven continuum manipulators with irregular geometry," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 2418–2428, 2021. 5

[46] N. Stefanakis, D. Pavlidi, and A. Mouchtaris, "Perpendicular cross-spectra fusion for sound source localization with a planar microphone array," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 9, pp. 1821–1835, 2017. 6