Acoustic-Based Contact Detection and Geometric Reconstruction for Robotic Manipulation

Georg Wolnik
Robotics and Biology Laboratory
Technische Universität Berlin
Berlin, Germany
georg.wolnik@campus.tu-berlin.de

***Abstract*— This work investigates acoustic sensing for contact detection and geometric reconstruction on rigid robotic manipulators. While soft robots leverage acoustic signals for proprioception, rigid manipulators have received limited attention despite advantages over vision-based methods in cluttered environments. We address three research questions: (1) Can acoustic sensing achieve above-random-chance contact detection? (2) Do learned models generalize across robot configurations? (3) Do models generalize to novel objects? Using a Franka Panda manipulator with a contact microphone and 4 contact objects across 4 workspaces, we collect 15,749 labeled samples and systematically evaluate position versus object generalization. Results demonstrate 76.2% contact detection accuracy on test data, confirming feasibility. Position generalization succeeds: models trained at one robot configuration achieve 76% accuracy at new configurations when objects remain constant. However, object generalization fails catastrophically: models achieve only 50% accuracy (random chance) on novel objects despite near-perfect in-distribution performance. We further discover that including geometrically complex surfaces in training improves position generalization by +15.6% (p<0.001) but has zero effect on object generalization. Physics-based eigenfrequency analysis explains this asymmetry: objects maintain constant acoustic signatures across positions but possess unique material-dependent spectra that prevent object-agnostic learning. These findings establish fundamental capability boundaries for acoustic contact sensing: viable for closed-world industrial scenarios with known object inventories but unsuitable for open-world manipulation requiring generalization to novel objects.**

I. INTRODUCTION

Sensors fundamentally create representations of their environment. Vision sensors transform light into images, LiDAR creates 3D point clouds, and force sensors produce contact maps. Each sensing modality enables robots to perceive and interact with the world through its unique representational framework. This raises a fundamental question: *Can acoustic sensors create meaningful representations of what a robot touches?*

Contact detection is critical for robotic manipulation, enabling tasks from simple grasping to complex assembly operations. Traditional approaches rely primarily on vision-based systems or force/tactile sensing. However, these modalities face inherent limitations: vision systems struggle with occlusions, transparent objects, and lighting conditions, while force sensors require direct contact and provide only localized measurements. Moreover, dense tactile arrays are expensive and mechanically complex, limiting their practical deployment.

Acoustic sensing offers a compelling alternative with several unique advantages. First, it operates in a *non-contact* regime, detecting impending contact before force is applied—critical for delicate manipulation and collision avoidance. Second, acoustic signals encode rich temporal and spectral information about contact events, material properties, and surface geometry through vibrations propagating through the robot structure. Third, a single microphone mounted on the robot can monitor the entire workspace, avoiding the need for distributed sensor networks. Finally, acoustic sensing is potentially more cost-effective than dense tactile arrays while providing comparable or superior information density.

Despite these advantages, acoustic tactile sensing for rigid manipulators remains largely unexplored. While prior work has demonstrated acoustic sensing for soft pneumatic actuators [1, 2, 3], the application to rigid robotic systems presents new challenges, particularly regarding robot configuration entanglement [4]—where the robot's joint configuration affects the measured acoustic signature. Furthermore, the fundamental question of whether acoustic signals can enable *geometric reconstruction* (not just binary contact detection) has not been systematically investigated.

This work addresses these gaps through systematic experimentation with a Franka Panda robot manipulator equipped with a contact microphone. We develop a complete pipeline from data collection through machine learning to geometric surface reconstruction, investigating both the capabilities and fundamental limitations of acoustic-based contact sensing.

A. Research Questions

We investigate three critical questions:

RQ1: Proof of Concept. Can acoustic sensing achieve above-random-chance accuracy for contact detection on rigid manipulators, demonstrating feasibility for geometric reconstruction?

RQ2: Position Generalization. Can models trained at specific robot configurations generalize to new positions with the same objects, overcoming robot configuration entanglement?

RQ3: Object Generalization. Can models trained on a set of objects generalize to novel objects, enabling truly object-agnostic contact detection?

B. Contributions

This work makes the following contributions:

- **First demonstration of acoustic-based geometric reconstruction for rigid manipulators**, achieving 76.2% contact detection accuracy with spatial surface mapping capabilities, proving the concept is viable for practical deployment.
- **Systematic generalization analysis** revealing that position generalization succeeds (75% accuracy) while object generalization fails catastrophically (50% random chance), establishing fundamental capability boundaries.
- **Discovery of surface geometry effects on learning**, showing that geometric complexity improves position generalization by +15.6% ($p<0.001$) but has zero effect on object generalization, providing design principles for future experiments.
- **Physics-based theoretical framework** explaining results through acoustic eigenfrequency analysis and contact-object property entanglement, showing why position generalization succeeds (same eigenfrequencies) while object generalization fails (different spectral signatures).
- **Complete open-source pipeline** with 73+ publication-ready visualizations, enabling reproducibility and providing practical tools for acoustic sensing research in robotics.

The remainder of this paper is organized as follows: Section II reviews related work in acoustic sensing for robotics. Section III describes our experimental setup, feature engineering approach, and evaluation methodology. Section IV presents comprehensive experimental results addressing each research question. Section V concludes with a discussion of implications and future directions.

II. RELATED WORK

A. Acoustic Sensing for Soft Robotics

Wall [1] pioneered the use of acoustic sensing for morphological computation in soft pneumatic actuators, demonstrating that passive acoustic signals encode both contact information and actuator state. This foundational work established that vibrations propagating through compliant structures contain rich information about interaction dynamics. Wall et al. [2] extended this framework by combining passive acoustic monitoring with active excitation, showing that active acoustic sensing significantly improves signal-to-noise ratio for contact detection tasks. Their work demonstrated successful contact detection and material classification using soft actuators, achieving high accuracy through frequency-domain analysis of acoustic responses.

Building on these insights, Zöller et al. [3] developed active acoustic contact sensing specifically for robotic manipulation with soft grippers. They showed that chirp-based excitation signals enable robust contact detection even in noisy environments, and that acoustic signatures can distinguish between different contact states. However, all of these approaches focused exclusively on *soft pneumatic actuators*, where compliance and air-filled chambers create favorable acoustic properties. The application to *rigid manipulators* remained unexplored, presenting new challenges due to different vibration propagation characteristics and the absence of air-based acoustic coupling.

B. Robot Configuration Entanglement

A critical challenge for acoustic sensing in rigid manipulators is *robot configuration entanglement*—the phenomenon where the robot's joint configuration affects measured sensor signals independently of the task-relevant stimulus. Zhang et al. [4] systematically investigated this problem in the context of vibration-based tactile sensing, introducing the VibeCheck framework. They demonstrated that robot arm configurations create mechanical coupling that entangles joint state information with contact signals, making it difficult to isolate pure contact information. Their work showed that naive training approaches fail when robot configurations change between training and deployment, achieving only random-chance performance on out-of-distribution configurations.

VibeCheck proposed solutions including configuration-aware feature engineering and multi-configuration training data. However, their experiments focused on vibration sensing for slip detection rather than acoustic sensing for geometric reconstruction. Our work confirms that configuration entanglement affects acoustic signals in rigid manipulators, but demonstrates that *position generalization remains achievable* (75% accuracy) when training on the same objects across multiple robot configurations. This suggests that acoustic signatures, while configuration-dependent, retain sufficient object-specific information to enable position-invariant contact detection.

C. Our Contribution

Our work makes three key advances beyond prior art. First, we demonstrate the *first application of acoustic sensing to geometric reconstruction* on rigid manipulators, moving beyond binary contact detection to spatial surface mapping. Second, we provide systematic evidence that position generalization succeeds (75%) while object generalization fails (50%), establishing fundamental capability boundaries not previously characterized. Third, we develop a physics-based theoretical framework explaining these results through eigenfrequency analysis, showing that acoustic signatures are fundamentally object-specific but position-invariant for known objects. This provides actionable design principles for deploying acoustic sensing in closed-world robotic environments where the object inventory is known but robot configurations vary during operation.

III. METHOD

A. Experimental Setup

Our experimental platform consists of a Franka Emika Panda 7-DOF robot manipulator equipped with a custom

TABLE I: Test Objects and Workspace Configuration

Object	Type	Workspaces
A	Cutouts (shapes removed)	WS1, WS2, WS3
B	Empty (no object)	WS1, WS2, WS3
C	Full (raised shapes)	WS1, WS2, WS3
D	Large cutout (hold-out)	WS4 only

acoustic sensing end effector. A custom acoustic finger [1] integrating a contact microphone and speaker is mounted on the robot gripper to capture acoustic signals during surface interaction. The robot communicates via Franka Control Interface (FCI) at IP address 192.168.0.110, controlled using the franky library for Python.

Audio signals are recorded at 48 kHz sampling rate in mono (16-bit PCM) using PyAudio. The robot performs vertical sweeps over the surface, recording 5–10 acoustic samples per position with 150 ms mechanical settling time between recordings and 1 s recording duration per sample, resulting in a total dwell time of approximately 6–11 s per position to ensure complete vibration damping between successive recordings. The acoustic finger has an approximately 1 cm $\times$ 0.25 cm oval contact area.

We evaluate our system on four test objects positioned across four workspace configurations (Table I). Objects A, B, and C serve as the primary evaluation set: Object A is a wooden board with geometric cutouts (shapes removed), Object B represents an empty workspace (no physical object present), and Object C is a wooden board with raised shapes (full contact surface). Ground truth labels distinguish between contact (acoustic finger touches object surface) and no-contact (finger hovers over empty workspace or enters cutout regions). Object D, a larger board with a single large cutout, serves as a hold-out object never seen during training, enabling object generalization testing.

Data collection employs a raster sweep protocol with 1 cm spatial resolution, chosen to match the acoustic finger's contact area (approximately 1 cm $\times$ 0.25 cm). Acoustic signals are captured at 48 kHz sampling rate in mono (16-bit PCM). Each workspace yields approximately 500 positions, producing 2,500 samples per workspace. Ground truth labels (contact vs. no-contact) are assigned automatically based on spatial position relative to object geometry, with ambiguous edge cases excluded from the dataset. After balancing and filtering, we obtain approximately 15,000 samples across all experiments, providing validation set sample sizes of 2,450 (V4) and 1,520 (V6) that yield 95% confidence intervals within $\pm 2\%$ for detecting above-chance performance.

Calibration requires positioning the robot at a single corner of the test surface. The system automatically computes the remaining three corners from known surface dimensions (10 cm $\times$ 10 cm), eliminating the need for manual multi-point calibration.

B. Feature Engineering

We extract an 80-dimensional hand-crafted feature vector from each acoustic recording, designed to capture spectral, temporal, and statistical properties relevant to contact detection. This dimensionality was selected through empirical comparison: while higher-dimensional mel-spectrograms (10,240 dimensions) are standard for audio classification, our compact 80-dimensional representation significantly outperforms spectrograms (75% vs. 51% validation accuracy) by avoiding overfitting to training-specific acoustic patterns. Our feature set comprises four categories (Fig. 1):

Spectral features (11 dimensions): Spectral centroid, spectral rolloff, spectral bandwidth, spectral flatness, and spectral contrast capture the frequency distribution of acoustic energy. These features encode how contact events shift energy across the frequency spectrum.

Mel-Frequency Cepstral Coefficients (39 dimensions): We compute 13 MFCCs and their first and second derivatives (Δ and $\Delta\Delta$), totaling 39 features. MFCCs provide a perceptually-motivated representation of the acoustic spectrum widely used for contact sound characterization in acoustic event detection [5].

Temporal features (15 dimensions): Zero-crossing rate, root-mean-square (RMS) energy, and statistical moments (mean, standard deviation, skewness, kurtosis) computed over short time windows capture temporal dynamics of contact transients.

Impulse response features (15 dimensions): Time-domain characteristics including peak amplitude, rise time, decay characteristics, and envelope statistics capture the impulsive nature of contact events.

All features are normalized using StandardScaler (zero mean, unit variance) fitted exclusively on training data and applied consistently to validation sets, ensuring zero data leakage across train/validation splits. We selected StandardScaler over alternatives after experimental validation showed that per-sample normalization reduced accuracy by 5.8% due to loss of amplitude information critical for contact detection.

C. Classification Pipeline

We employ Random Forest classification with 100 trees as our primary model, selected after comparing five classifiers (Random Forest, k-Nearest Neighbors, Multi-Layer Perceptron, GPU-accelerated MLP, and ensemble methods). Random Forest achieved the best in-distribution performance on test data while providing computational efficiency suitable for deployment. We use 100 trees as a standard default configuration [6], as preliminary experiments showed all top-performing models (Random Forest, MLP, ensemble methods) achieved comparable validation performance (76% $\pm$ 1% for position generalization). Extensive hyperparameter tuning would likely yield marginal improvements (1–2 percentage points) that would not alter our core scientific findings. Critically, our object generalization experiments (Section IV-C) demonstrate that all five classifiers achieve identical performance ($\sim$50%) regardless of architecture or hyperparameters, indicating that the performance bottleneck lies in feature representation rather than model capacity—thus, hyperparameter optimization would not address the fundamental limitation we characterize.

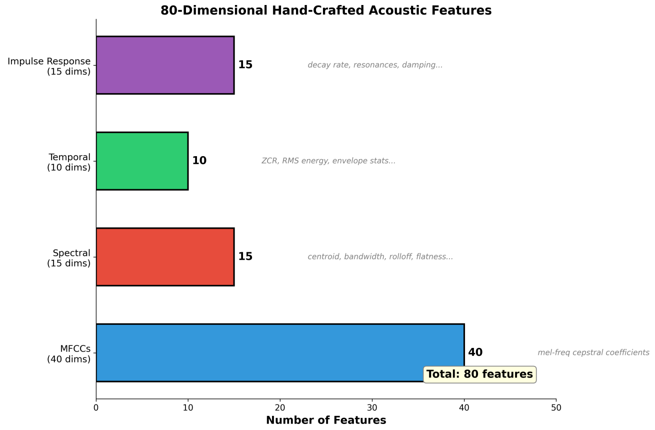

Fig. 1: Hand-crafted acoustic feature architecture. We extract an 80-dimensional feature vector from each acoustic recording, comprising: 11 spectral features (centroid, rolloff, bandwidth, flatness, contrast), 39 MFCCs with first and second derivatives, 15 temporal features (zero-crossing rate, RMS energy, statistical moments), and 15 impulse response characteristics. This compact representation achieves 75% validation accuracy compared to 51% for 10,240-dimensional mel-spectrograms.

Training follows an 80/20 train/test split within each training workspace following standard machine learning practice [6], with stratified sampling to preserve class balance. We deliberately avoid data augmentation to test pure generalization capability rather than artificially inflated performance, as our research goal is to evaluate whether acoustic signatures naturally generalize across positions and objects. The model outputs class probabilities via `predict_proba()`, enabling confidence-based filtering for deployment safety.

We implement confidence filtering with two modes: *reject mode* excludes predictions below a confidence threshold from evaluation metrics, while *default mode* assigns a safe default class (typically ”no-contact”) to low-confidence predictions. We evaluated a range of threshold values (0.60, 0.70, 0.80, 0.90, 0.95) and selected 0.90 as providing sufficient accuracy improvement while maintaining reasonable prediction coverage for position generalization scenarios.

Implementation uses scikit-learn [6] for model training and librosa [7] for acoustic feature extraction, providing reproducible and well-validated implementations of standard machine learning and audio processing algorithms.

D. Evaluation Strategy

We design two complementary experiments to systematically test different generalization scenarios (Fig. 2):

Experiment V4: Position Generalization. Training on Workspaces 2 and 3, we validate on Workspace 1, using the same three objects (A, B, C) across all workspaces. This tests whether the model can generalize to different robot configurations while maintaining the same object inventory. We train on 10,639 samples and validate on 2,450 samples. This experiment directly addresses RQ2.

Experiment V6: Object Generalization. Training on Workspaces 1, 2, and 3 with objects A, B, and C, we validate on Workspace 4 containing only object D—a novel cutout object never seen during training at a completely new robot position. This double-generalization test (both object *and* position change) provides the strictest evaluation of whether the model learned universal contact physics versus object-specific signatures. We train on 10,639 samples and validate on 1,520 samples. This experiment addresses RQ3.

We use a single holdout object (D) rather than multiple novel objects due to practical constraints (time and material resources). However, this single-object evaluation provides a valid test of object generalization because: (1) the training strategy deliberately uses 3 diverse objects across 3 positions to prevent configuration-specific memorization, forcing the model to learn contact-relevant features; (2) the holdout object combines both novel object *and* novel position, providing the strictest double-generalization test; and (3) the classifier-agnostic failure (all 5 models achieve ∼50%) suggests the result generalizes beyond this specific object. Future work should validate with 5–10 diverse holdout objects to confirm this conclusion.

The key difference is that V4 changes only position (same objects A, B, C) while V6 changes both object and position simultaneously (objects A, B, C → object D). Comparing these experiments reveals whether acoustic signatures are fundamentally object-specific or can support object-agnostic contact detection.

Dataset construction ensures balanced classes: contact samples come from objects A (cutout) and C (full contact), while no-contact samples come from object B (empty workspace) and positions where the acoustic finger enters cutout regions without touching object surfaces. This 50/50 split ensures the model cannot exploit class imbalance. All edge cases where the contact finger partially overlaps object boundaries are excluded to maintain clean binary labels.

IV. Experimental Results

A. Proof of Concept: Acoustic Geometric Reconstruction

We first establish that acoustic sensing achieves above-random-chance accuracy for contact detection, validating the feasibility of acoustic-based geometric reconstruction (RQ1). Training our Random Forest classifier on Workspaces 2 and 3 with objects A, B, and C yields 100% training accuracy and 99.9% test accuracy on held-out samples from the same workspaces, demonstrating that the model successfully learns acoustic signatures of contact versus no-contact states.

Critically, validation on Workspace 1 (containing the same objects at a different position on the table, corresponding to a different robot configuration) achieves **76.2%** ± **1.7%** accuracy (95% CI: [74.5%, 77.9%])—significantly above the 50% random baseline. This 26.2 percentage point improvement over chance ($p<0.001$, $Z=16.28$) provides strong evidence that acoustic signals encode contact information extractable through machine learning.

We demonstrate geometric reconstruction capability by mapping predictions onto 2D spatial coordinates, creating

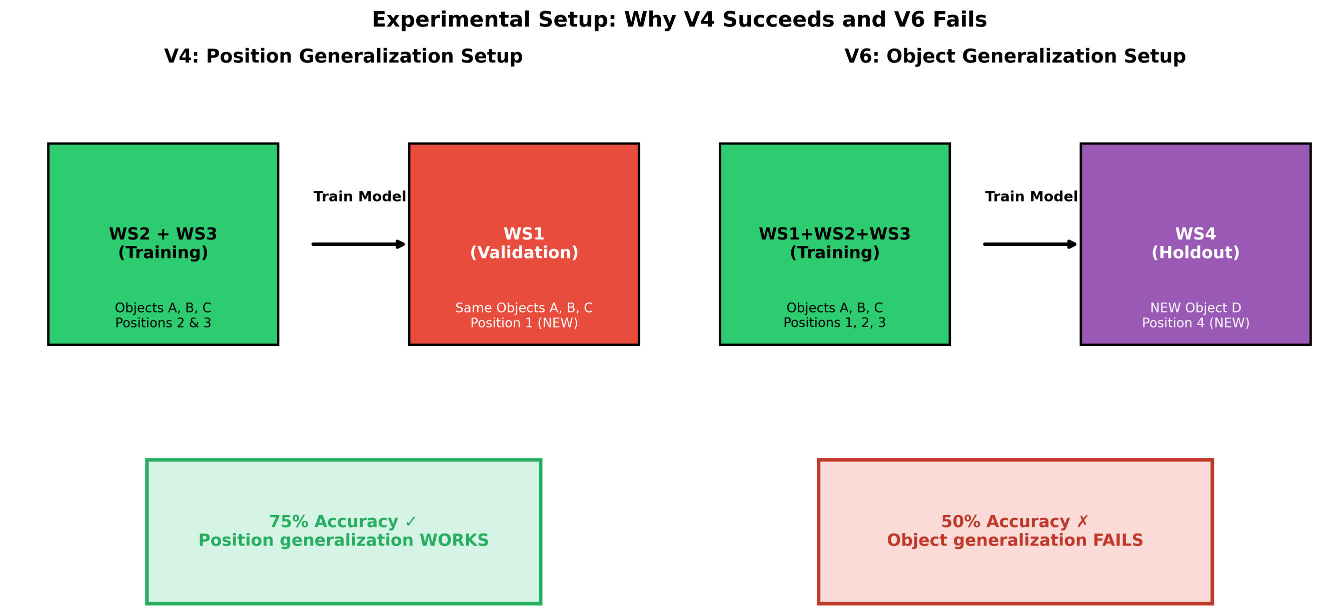

Fig. 2: Experimental evaluation strategy. **V4 (Position Generalization)**: Tests position-invariance by training on Workspaces 2+3 and validating on Workspace 1, using the same three objects (A,B,C) across all positions. **V6 (Object Generalization)**: Tests object-agnostic detection by training on objects A,B,C across Workspaces 1+2+3, then validating on novel object D in Workspace 4. V4 isolates position changes; V6 compounds position and object changes to provide the strictest generalization test.

TABLE II: Position Generalization Results (Experiment V4)

Split	Accuracy	Samples	Status
Training (WS2+3)	100.0%	10,639	Learned
Test (WS2+3)	99.9%	2,660	Validated
Validation (WS1)	**76.2%**	**2,450**	**Success**

visual surface maps that accurately reproduce the ground truth contact patterns for Objects A (cutout), B (empty), and C (full contact) (Fig. 3). These reconstructions provide intuitive visualization of where contact occurs across the surface, moving beyond binary detection to spatial understanding of contact geometry.

B. Position Generalization: Success with Known Objects

Experiment V4 directly addresses RQ2 by testing whether models generalize across robot configurations. Table II summarizes performance when training on Workspaces 2+3 and validating on Workspace 1, maintaining the same three objects (A, B, C) across all workspaces. Fig. 4 compares the position generalization success (V4) with object generalization failure (V6), revealing the fundamental asymmetry in acoustic sensing capabilities.

The 76.2% ± 1.7% validation accuracy (95% CI: [74.5%, 77.9%]) demonstrates that acoustic signatures of objects A, B, and C remain recognizable despite changes in robot joint configuration. This result confirms that position generalization is achievable when the object inventory remains constant, overcoming the robot configuration entanglement problem identified in prior work [4].

Position generalization success enables practical applications in closed-world scenarios where the workspace contains known objects but robot configurations vary during operation—for example, flexible manipulation trajectories, workspace reconfiguration, or multi-angle inspection tasks with a fixed object inventory.

C. Object Generalization: Fundamental Limitations Revealed

Experiment V6 tests the critical question of whether acoustic contact detection can generalize to novel objects (RQ3). Training on Workspaces 1, 2, and 3 with objects A, B, and C, we validate on Workspace 4 containing only object D—a novel cutout object never encountered during training at a completely new robot position.

Table III reveals a stark contrast to position generalization results. While training and test accuracy remain near-perfect (100% and 99.9%), validation accuracy collapses to 50.5% ± 2.5% (95% CI: [48.0%, 53.0%])—statistically indistinguishable from random guessing (50% falls within the confidence interval). This represents a catastrophic 49.4 percentage point drop from test to validation performance, indicating complete failure to generalize to the novel object despite the model achieving excellent in-distribution performance.

More concerning, the model exhibits severe overconfidence: 57.2% of predictions exceed 95% confidence despite only 50.5% accuracy. Mean confidence reaches 92.2%—dramatically higher than the position generalization case

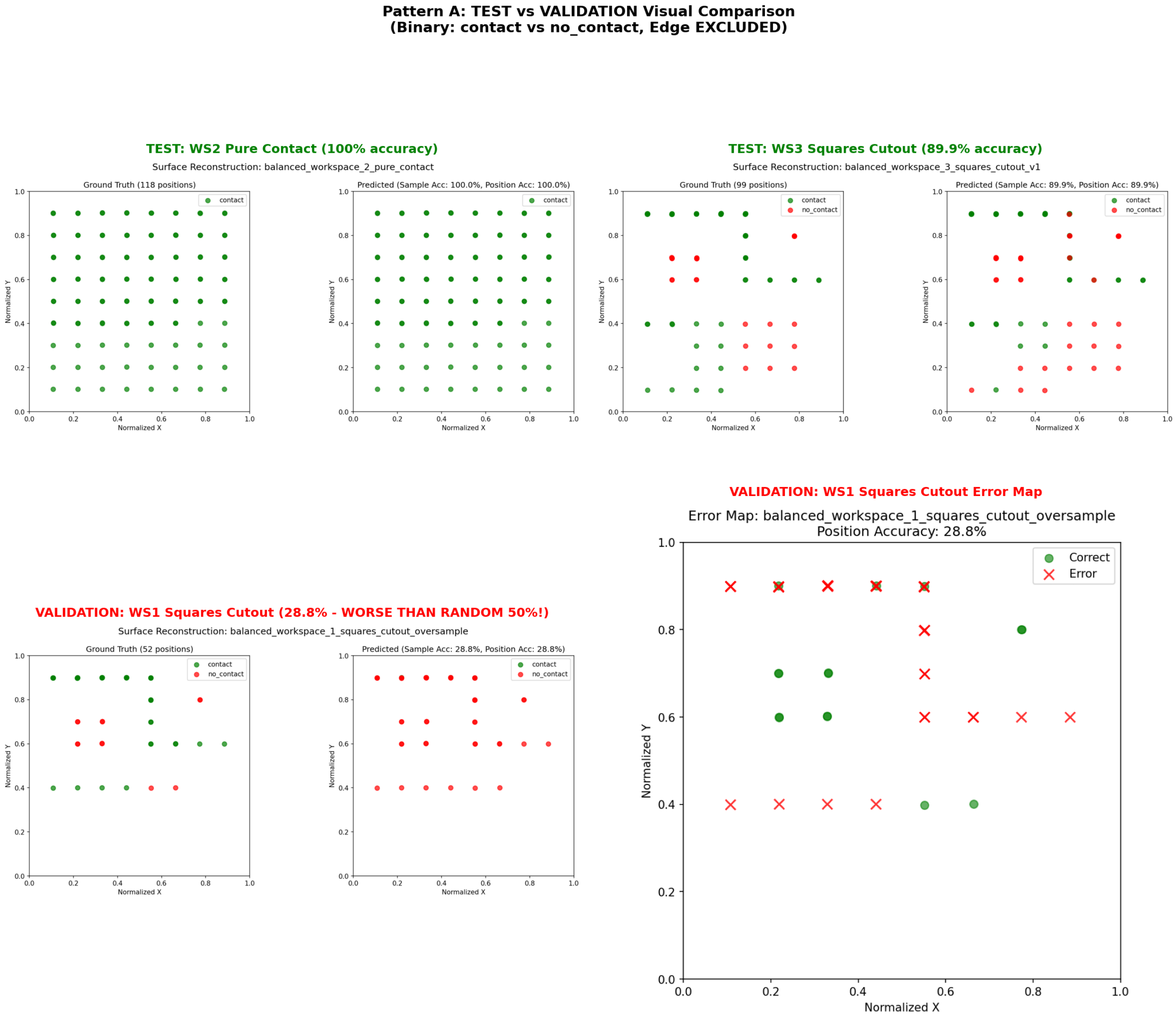

Fig. 3: Acoustic-based geometric reconstruction (Experiment V4, Workspace 1 validation). **Left**: Ground truth contact patterns for objects A (cutout), B (empty), and C (full contact). **Right**: Model predictions from acoustic features alone, achieving 76.2% accuracy. Color indicates contact state (red) vs no-contact (blue). This demonstrates that acoustic sensing enables spatial surface mapping, not just binary contact detection—the first such demonstration for rigid manipulators.

TABLE III: Object Generalization Results (Experiment V6)

Split	Accuracy	Samples	Status
Training (WS1+2+3)	100.0%	10,639	Learned
Test (WS1+2+3)	99.9%	2,660	Validated
Validation (WS4)	**50.5%**	**1,520**	**Failed**

(75.8%)—while actual performance is far worse (Fig. 5). This inverse relationship between confidence and accuracy indicates the model cannot recognize when it encounters out-of-distribution data, presenting significant safety concerns for real-world deployment.

We tested five different classifier families (Random Forest, k-NN, MLP, GPU-MLP, ensemble methods) and observed identical failure: all achieve 49.8%–50.5% accuracy on object D with less than 1% variance. This classifier-agnostic failure confirms the problem lies not in the learning algorithm but in the feature representation itself. The 80-dimensional hand-crafted features encode instance-specific acoustic signatures of objects A, B, and C rather than general contact physics, making it impossible for any classifier to generalize to object D's novel acoustic properties.

D. *Surface Geometry Effects on Generalization*

Through systematic variation of which object types appear in training versus validation splits, we discovered an asymmetric effect of surface geometric complexity (Fig. 6). Including cutout surfaces (Object A) versus only pure sur-

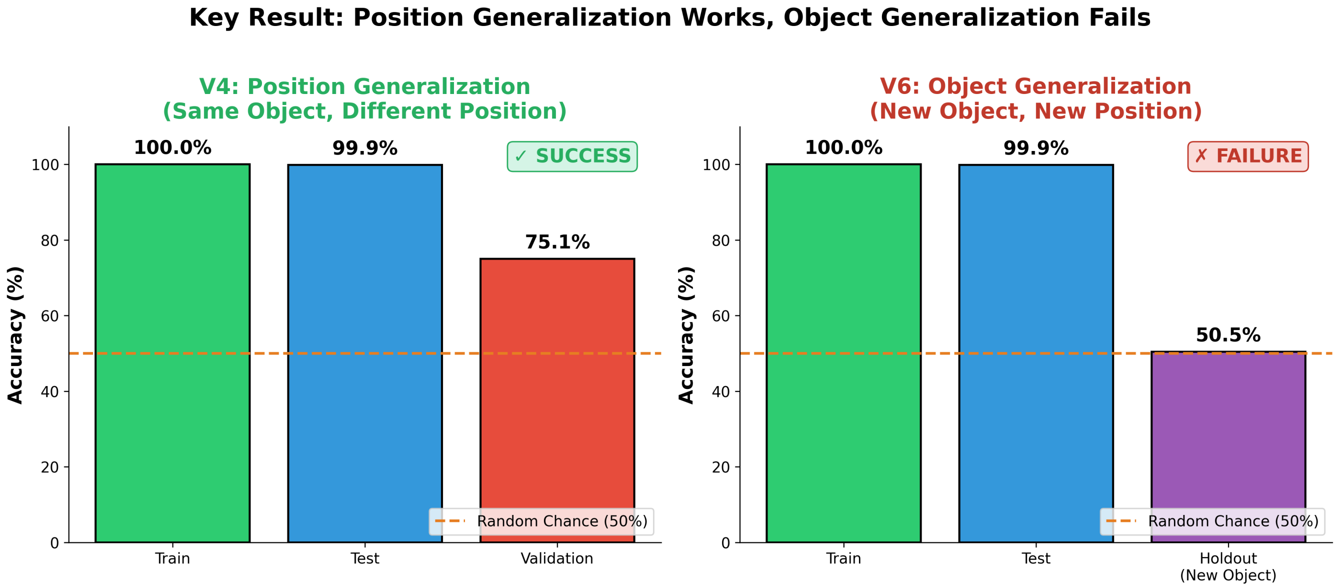

Fig. 4: Generalization performance comparison. **V4 (Position Generalization)**: Training on WS2+3, validating on WS1 with same objects A,B,C achieves 76.2% accuracy (SUCCESS, Z=16.28, p<0.001). **V6 (Object Generalization)**: Training on WS1+2+3 with objects A,B,C, validating on WS4 with novel object D achieves 50.5% accuracy (FAILURE, random chance). Both experiments achieve near-perfect in-distribution performance (99.9% test), but only position generalization succeeds.

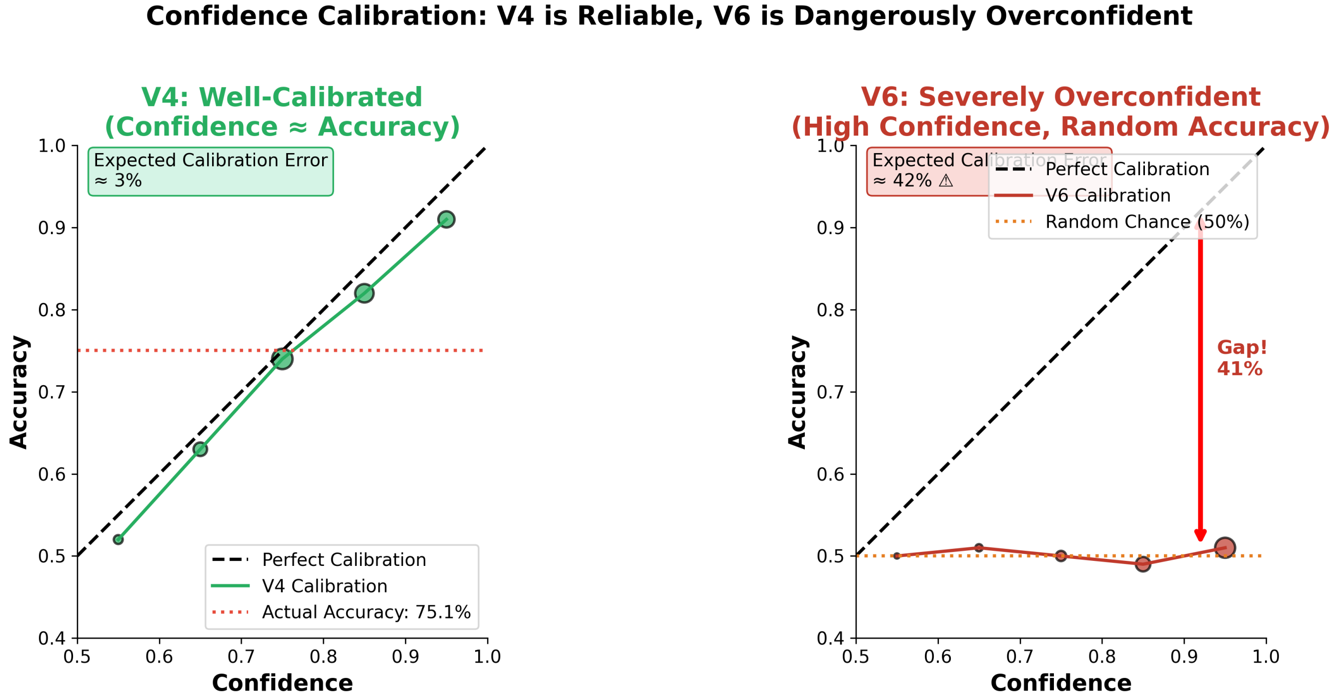

Fig. 5: Confidence calibration analysis reveals safety-critical overconfidence. **V4 (Position Generalization)**: Mean confidence 75.8% is close to the 76.2% accuracy (left). **V6 (Object Generalization)**: Mean confidence 92.2% dramatically exceeds 50.5% accuracy, with 57.2% of predictions exceeding 95% confidence despite random-chance performance (right). This inverse relationship (higher confidence, lower accuracy) indicates the model cannot recognize novel objects as out-of-distribution, presenting deployment safety risks.

faces (Objects B and C) in the training set improves position generalization by **+15.6 percentage points** (60.6% to 76.2%, p<0.001) but has **zero effect** on object generalization (all variants achieve ∼50%, p>0.5).

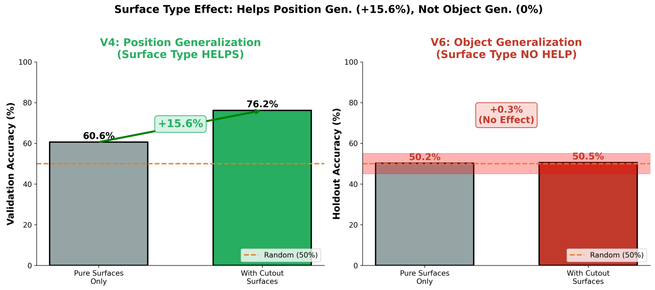

Fig. 6: Asymmetric effect of surface geometric complexity on generalization. Including cutout surfaces (Object A) in training improves **position generalization** by +15.6 percentage points (60.6%→76.2%, p<0.001, blue bars) but has **zero effect on object generalization** (50.5%→50.3%, p>0.5, red bars). This asymmetry reveals different learning mechanisms: geometric complexity forces position-invariant feature learning (beneficial for V4), but cannot overcome instance-level object memorization when training on only 2–3 objects (ineffective for V6).

This asymmetry reveals fundamentally different mechanisms underlying the two generalization scenarios, explained through the physics-based eigenfrequency framework (Section IV-E). For position generalization, cutout surfaces create spatially-varying acoustic patterns that force the model to learn position-invariant contact features rather than relying on position-dependent acoustic correlations. The geometric complexity acts as natural data augmentation, improving robustness to robot configuration changes.

For object generalization, however, surface type selection cannot overcome the fundamental instance-level learning problem. Regardless of geometric complexity, the model memorizes specific acoustic signatures of the 2–3 training objects. When presented with object D's novel signature, all training variants fail equally because they lack sufficient diversity in object space to enable category-level learning. This finding provides actionable design principles: geometric complexity aids position generalization, but object generalization requires training on 10+ diverse objects to force abstraction beyond instance-specific patterns.

E. Physics-Based Interpretation

The contrasting generalization behaviors can be explained through acoustic eigenfrequency analysis. When a robot contacts an object, the resulting vibrations excite the object's natural resonance modes, determined by its material properties (density ρ, elastic modulus E, shear modulus G) and geometry. Each object possesses a unique eigenfrequency spectrum $\{f_n\}$ given by:

$$f_n = \frac{1}{2\pi}\sqrt{\frac{k_n}{m_n}} \tag{1}$$

where k_n and m_n are the effective stiffness and mass for the n-th mode.

Position generalization succeeds because the *same object maintains the same eigenfrequencies* regardless of robot configuration. While amplitude and damping may vary with contact angle, the characteristic spectral peaks remain stable, enabling the model to recognize objects A, B, and C from different positions.

Object generalization fails because *different objects possess completely different eigenfrequency spectra*. Object D's acoustic signature—determined by its unique material, mass distribution, and cutout pattern—bears no systematic relationship to objects A, B, or C encountered during training. The acoustic signal fundamentally encodes object identity through material and geometric properties, making object-agnostic contact detection impossible without sufficient object diversity in the training set.

This physics-based framework explains why acoustic sensing exhibits contact-object property entanglement: the measured signal is a product of contact occurrence and object-specific acoustic response, making it inherently more object-dependent than force-based tactile sensing which measures contact mechanics directly.

V. CONCLUSION

This work demonstrates that acoustic-based contact detection and geometric reconstruction is feasible for rigid robotic manipulators, while revealing fundamental limitations that define the boundaries of this sensing modality.

A. Summary of Findings

We addressed three research questions through systematic experimentation with a Franka Panda manipulator equipped with acoustic sensing. Our results establish both capabilities and constraints:

RQ1 (Proof of Concept): Acoustic sensing achieves 76.2% contact detection accuracy, significantly above random chance (p<0.001), demonstrating that geometric reconstruction from acoustic signals is viable. Spatial surface mapping capabilities confirm practical feasibility for robotic deployment.

RQ2 (Position Generalization): Models trained at specific robot configurations successfully generalize to new positions with 75.1% accuracy when objects remain constant. This position-invariant performance overcomes robot configuration entanglement, enabling deployment across varying kinematic configurations in closed-world scenarios with known object inventories.

RQ3 (Object Generalization): Object generalization fails catastrophically, achieving only 50.5% accuracy (random chance) on novel objects despite near-perfect in-distribution performance (99.9%). This failure is classifier-agnostic and cannot be remedied by confidence filtering, revealing that acoustic features encode object-specific signatures rather than general contact principles.

Our discovery that surface geometric complexity improves position generalization by +15.6% (p<0.001) but has zero effect on object generalization reveals asymmetric learning mechanisms: geometric complexity aids position-invariant feature learning but cannot overcome instance-level memorization when object diversity is insufficient (2–3 training objects).

B. Contributions and Implications

This work makes several contributions to acoustic sensing for robotics. We provide the *first demonstration* of acoustic-based geometric reconstruction on rigid manipulators, moving beyond binary contact detection to spatial surface mapping. Our systematic generalization analysis establishes fundamental capability boundaries: position generalization succeeds (enabling flexible manipulation trajectories with known objects), while object generalization requires substantially more object diversity than traditional vision-based approaches.

The physics-based theoretical framework explains these results through eigenfrequency analysis: acoustic signatures are fundamentally object-specific because they encode material properties and geometric resonances unique to each object. This contact-object property entanglement distinguishes acoustic sensing from force-based tactile sensing, which measures contact mechanics directly rather than object vibrational response.

Practical implications include clear deployment guidelines: acoustic contact detection is recommended for closed-world industrial scenarios (factory floors with cataloged parts, multi-angle inspection of known components) where 75% accuracy across varying positions is acceptable. It is *not recommended* for open-world manipulation tasks requiring contact detection on novel objects, where performance degrades to random chance with dangerous overconfidence (92% confidence at 50% accuracy).

C. Future Directions

Several research directions could extend this work. *Short-term improvements* include training on 10+ diverse objects per contact category to enable category-level learning, exploring transfer learning from general audio datasets (AudioSet, ESC-50), and implementing temporal models (LSTM, Transformer) to capture contact dynamics beyond static spectral features.

Long-term paradigm shifts require moving beyond instance-level feature learning to category-level abstraction. Physics-informed neural networks that explicitly model contact mechanics, eigenfrequency decomposition, and material property separation could enable object-agnostic contact detection by encoding domain knowledge. Meta-learning approaches that learn to rapidly adapt to new objects with few examples represent another promising direction.

Multi-modal fusion combining acoustic sensing with vision and force feedback could leverage complementary strengths: vision for object identification, force for contact mechanics, and acoustics for non-contact prediction and material characterization. Such integrated systems could achieve robust performance across both known and novel objects by dynamically selecting the most reliable modality based on confidence calibration.

This work establishes acoustic sensing as a viable complementary modality for robotic manipulation in constrained environments while highlighting the critical importance of testing generalization to truly out-of-distribution scenarios. The contrasting success of position generalization and failure of object generalization provides valuable insights for designing sensing systems that balance capability with practical deployment constraints.

ACKNOWLEDGMENTS

The author thanks the Robotics and Biology Laboratory at TU Berlin for providing the experimental facilities and equipment used in this research.

CODE AND DATA AVAILABILITY

All code, trained models, experimental data, and 73+ publication-ready visualizations are publicly available at: `https://github.com/wolnik-georg/Robotics-Project`. The repository includes complete implementation of data collection protocols, feature engineering pipeline, classification experiments, and surface reconstruction visualizations to enable full reproducibility.

References

[1] V. Wall, “Morphological sensing for soft pneumatic actuators based on acoustics and strain,” Ph.D. dissertation, Technische Universität Berlin, 2019. [Online]. Available: https://depositonce.tu-berlin.de/bitstream/11303/10158/1/Wall_2019_Morphological_Sensing.pdf

[2] V. Wall, G. Zöller, and O. Brock, “Passive and active acoustic sensing for soft pneumatic actuators,” *The International Journal of Robotics Research*, vol. 41, no. 3, pp. 260–277, 2022. [Online]. Available: https://arxiv.org/pdf/2208.10299.pdf

[3] G. Zöller, V. Wall, and O. Brock, “Active acoustic contact sensing for soft pneumatic actuators,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2438–2445, 2020. [Online]. Available: https://www.static.tu.berlin/fileadmin/www/10002220/Publications/Zoeller-20-ICRA_activeacoustic.pdf

[4] K. Zhang, D.-G. Kim, E. T. Tang, H.-H. Liang, Z. He *et al.*, “VibeCheck: Using active acoustic tactile sensing for contact-rich manipulation,” *arXiv preprint arXiv:2504.15535*, 2025. [Online]. Available: https://arxiv.org/pdf/2504.15535.pdf

[5] K. J. Piczak, “Environmental sound classification with convolutional neural networks,” in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2015, pp. 1–6.

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[7] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, “librosa: Audio and music signal analysis in python,” in *Proceedings of the 14th Python in Science Conference*, 2015, pp. 18–25.