Geometrical Reconstruction using Acoustic Tactile Sensing

Can a Robot Hear Whether It Touches Something?

Georg Wolnik

Robotics Project - Robotics and Biology Laboratory - TU Berlin

2nd February 2026

Motivation

The Problem

Vision isn't always available

- Robot arm/gripper occludes camera view
- Low-light, dusty, or special use cases
- Need for contact-based perception

Why Does This Matter?

- **Unexplored modality:** Acoustic sensing is underutilized in robotics
- **Cheap & simple:** Just a contact microphone
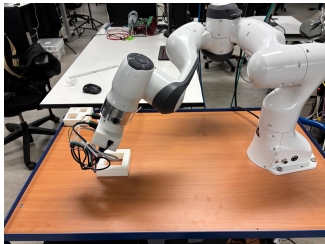- **Rich information:** Sound encodes material, geometry, contact dynamics

Project Goal

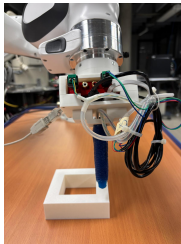*Can we reconstruct surface geometry from acoustic signals?*

- Detect contact events during robot motion
- Classify surface shapes acoustically
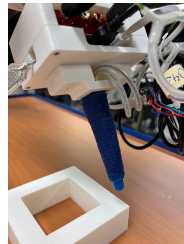- Evaluate generalization capabilities

Setup

Full Setup

Sensor Mounting

Close-up

Why Robot-Mounted Data Collection?
- ✔ **Controlled variables:** Consistent angle, pressure, and sweep velocity
- ✔ **Repeatability:** Systematic and reliable sweeps
- ✔ **Scalability:** Automated collection of datasets
- ✔ **Real-world relevance:** Directly applicable to robotic manipulation tasks

The Hidden Challenge: Robot Arm Affects the Signal

⚠ The Problem

The robot arm is part of the acoustic system
Sound waves travel through:
Surface → Probe → **Robot Arm** → Microphone

What This Means

- **Different arm positions** = Different acoustic responses
- Joint angles change wave propagation paths
- Same contact event sounds **different** at different positions

The Key Insight

Signal = Contact Information ⊗ Arm Configuration
The signal contains **both** — they are entangled!

(Also observed in VibeCheck [Zoller et al., 2024])

Why Multiple Workspaces?

WS1: Position A
Arm config θ_1

WS2: Position B
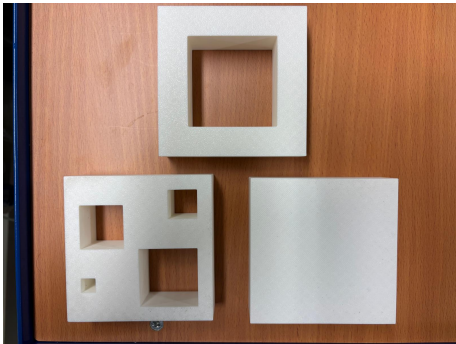Arm config θ_2

WS3: Position C
Arm config θ_3

Question:
Will a classifier trained at
one position work at another?

Data Collection: Objects and Workspaces

Test Objects

A: Cutouts — B: Empty — C: Full — D: Big
Cutout

Complete Dataset Overview

	Obj A Cutouts	Obj B Empty	Obj C Full	Obj D Big Cut
WS1	✓	✓	✓	
WS2	✓	✓	✓	
WS3	✓	✓	✓	
WS4				✓

Green: Proof of Concept data (A,B,C in WS1-3)

Red: Final Validation data (D in WS4)

?

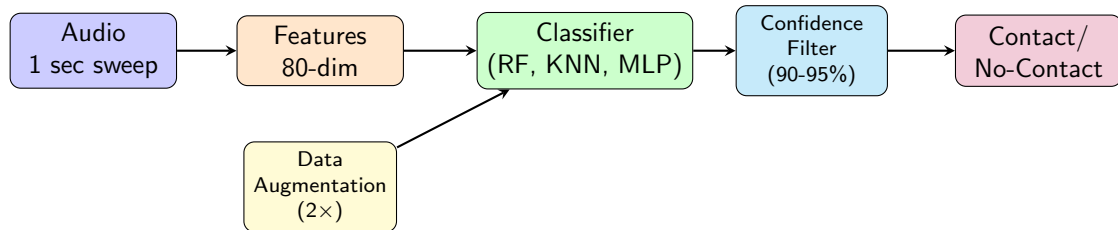# Can we reconstruct geometric shapes using binary surface classification?

Task	Target
Binary classification	Geometric reconstruction
Contact vs. No-Contact	**on data from the same distribution**

Method: Acoustic Feature Classification

80-Dimensional Features

- **MFCCs + Deltas** (39): Frequency content
- **Spectral** (11): Energy distribution
- **Temporal** (15): Contact dynamics
- **Impulse Response** (15): Resonance

Data Augmentation

- Noise injection
- Time shifting
- Pitch variation
- Gain scaling
- Time stretching

Why Hand-Crafted?

- Hand-crafted: **70%**
- Spectrograms: **59%**
- 80 vs 10,240 dims

Result: Excellent Classification on data from same distribution
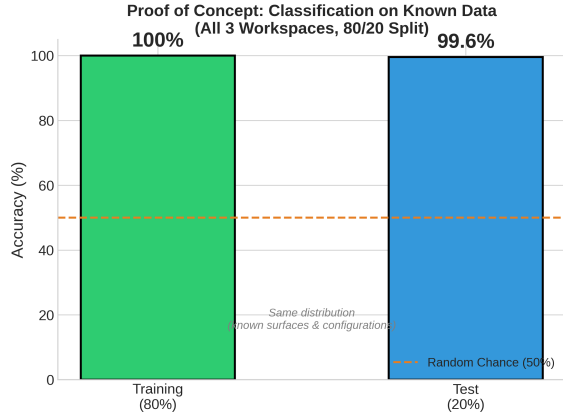

~**99%**
Test Accuracy
(Known surfaces & configurations)
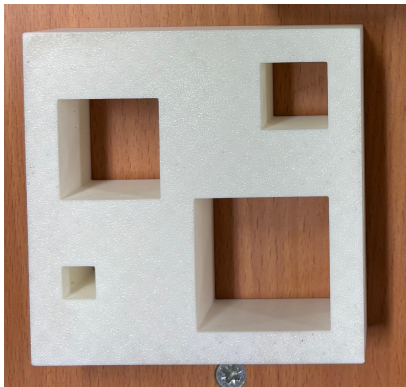
Dataset Split

3 Objects × 3 Workspaces
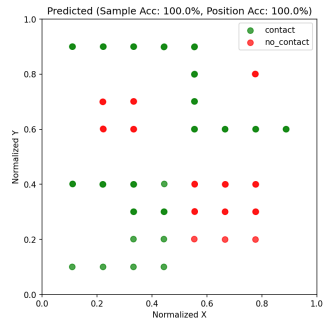
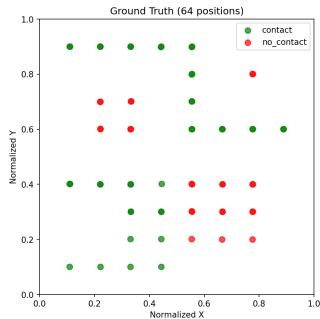Training: 80%

Test: 20%

Same distribution

Proof of Concept: Classification on Known Data
(All 3 Workspaces, 80/20 Split)

100%

99.6%

Same distribution
Known surfaces & configurations

Random Chance (50%)

Training
(80%)

Test
(20%)

Accuracy (%)

Reconstruction: Excellent Performance on Test Data

Object View

Ground Truth vs Prediction

Reconstruction: Full Contact Surface

Object View

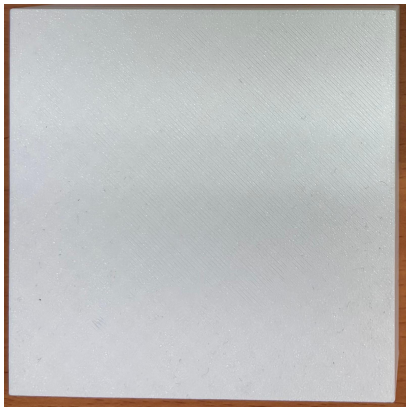## Ground Truth vs Prediction

Surface Reconstruction: balanced_workspace_2_pure_contact
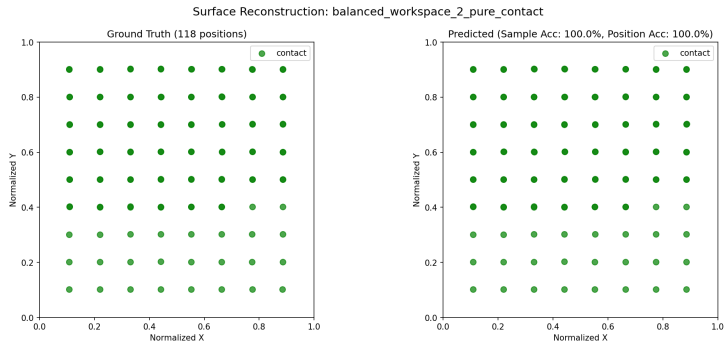
Reconstruction: No-Contact Surface

Ground Truth vs Prediction
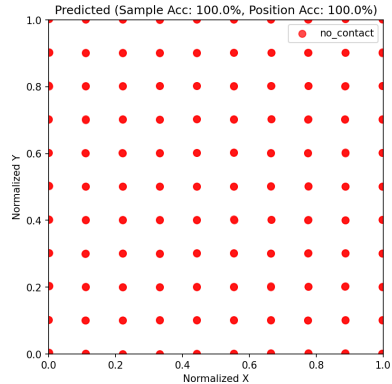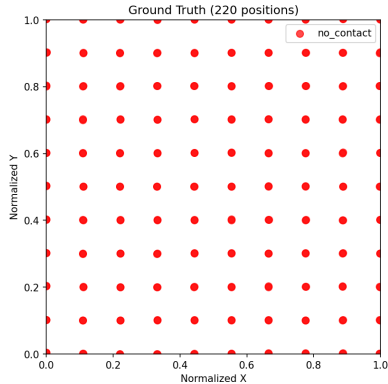
Surface Reconstruction: balanced_workspace_2_pure_no_contact

Now: Testing Generalization

✅ Proof of Concept Achieved!

Acoustic contact classification works on data from the same distribution

~100% test accuracy with excellent reconstruction

🔍
But does it generalize?

Two critical questions:

What if we test on a **new workspace**?	What if we test on a **new object**?
Train: WS1+WS3 → Test: WS2	Train: A,B,C → Test: D (+ new workspace)
Same objects, **different robot positions**	**Completely unseen object & workspace**
→ Configuration Generalization	**→ Object Generalization**

Challenge 1: Configuration Generalization

Experimental Setup

Training: WS1 + WS3

| WS1 | WS3 |

Same objects (A, B, C)

$\downarrow$

Validation: WS2 (unseen)

| WS2 |

Same objects, different robot configurations

Performance Results

Dataset	Accuracy
Training (WS1+WS3)	**100%**
Test (WS1+WS3, 20%)	**~100%**
Validation (WS2)	**65-70%**

Configuration Entanglement

65-70%

Better than random (50%), but significant drop

Signal = Contact $\otimes$ Configuration

Model partially learned configuration-specific features.
More diverse workspace data likely needed.

Challenge 2: Object Generalization

Object Generalization Setup

Training: 3 Objects (A, B, C)

| A: Cutouts | B: Empty | C: Full |

Train: WS1 + WS2 + WS3

↓

Hold-out: Object D

| D: Big Cutout |

Test: WS4 (new)

New object + new workspace
Complete generalization test

Performance Results

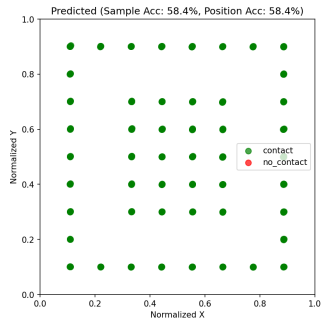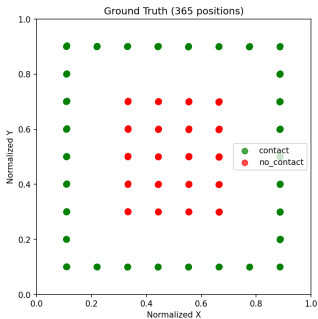| Dataset	Accuracy
Training (A,B,C on WS1-3)	**100%**
Test (A,B,C on WS1-3)	**∼100%**
Hold-out (D on WS4)	**50%**

Critical Discovery

50% = Random Chance

Model completely fails on unseen object

Reconstruction on new object + workspace fails

New Object + WS

Train: A-C → Test: D on WS4

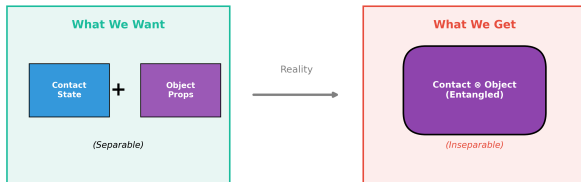Surface Reconstruction: balanced_holdout_oversample

50%

Understanding the Failure: The Entanglement Problem

The Entanglement Problem

Why Acoustic Signatures Cannot Separate Contact from Object Properties

What We Want

| Contact State | + | Object Props |

(Separable)

Reality →

What We Get

Contact ⊗ Object (Entangled)

(Inseparable)

Physical Reality:
- Acoustic signal S(t) = f(Contact, Object_Material, Object_Mass, Object_Geometry, ...)
  - These factors are multiplicatively coupled (⊗), not additively separable (+)
  - Model learns: "Object A sounds like THIS when touched" (instance-specific)
- Model cannot learn: "Contact sounds like THIS regardless of object" (category-level)

Result:
- Same object, different position → Features still correlate → 75% accuracy ✓
  - Different object → Completely different feature space → 50% accuracy ✗

The Core Problem

$$\text{Signal} = \text{Contact} \otimes \text{Object}$$

The acoustic signature contains **BOTH**:

1. Contact state information
2. Object identity information

Physics Explanation

- Each object has unique eigenfrequencies
- Material, size, geometry, mass affect acoustic response
- Model learned **"Object A contact"** not **"contact in general"**

Conclusions

✅ Main Achievement: Proof of Concept

"Acoustic-based geometric reconstruction IS POSSIBLE"
(on data from the same distribution: ∼100% test accuracy, excellent reconstruction)

What Works

- ∼**99%** test accuracy
  (known surfaces & configurations)
- Excellent surface reconstruction
- Hand-crafted features outperform spectrograms (70% vs 59%)
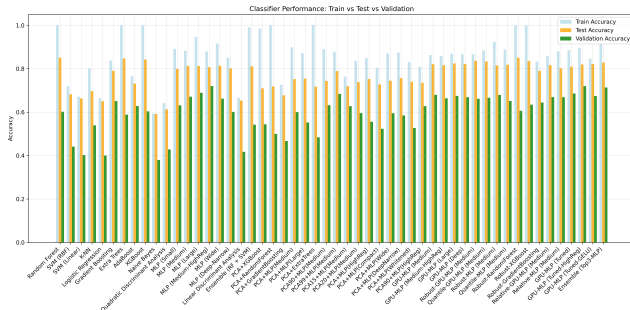- 128× fewer features, better results

Discovered Limitations

- **Configuration entanglement:**
  65-70% on unseen workspace
  Signal = Contact ⊗ Configuration
- **Object entanglement:**
  50% on unseen object (random)
  Signal = Contact ⊗ Object

Future Directions

- **Configuration invariance:** Collect data from many more diverse workspaces/positions
- **Object invariance:** Train on many varied objects to disentangle contact from object identity

Questions?

Backup: All Classifiers Comparison

Classifier Performance: Train vs Test vs Validation

Top Classifiers (Validation)

Classifier	Val Acc.
MLP (Medium-HighReg)	**72.0%**
GPU-MLP (Tuned-HighReg)	71.9%
Ensemble (Top3-MLP)	71.3%
MLP (Large)	68.9%
GPU-MLP (Tuned)	68.6%
Gradient Boosting	65.1%
Random Forest	60.1%

Key Finding

MLPs with regularization
outperform tree-based methods
on validation data

Backup: 80-Dimensional Feature Set

MFCCs + Deltas (39 dims)

- 13 MFCC coefficients
- 13 MFCC delta (velocity)
- 13 MFCC delta-delta (acceleration)

Spectral Features (11 dims)

- Spectral centroid
- Spectral bandwidth
- Spectral rolloff
- Spectral flatness
- Spectral contrast (7 bands)

Temporal Features (15 dims)

- Zero-crossing rate
- RMS energy
- Mean, std, skewness, kurtosis
- Onset strength
- Tempo estimation

Impulse Response (15 dims)

- Transfer function peaks
- Decay characteristics
- Resonance frequencies
- Damping coefficients

Backup: Object and Workspace Details

Training Objects (A, B, C)

- **Object A:** Cutouts
  Wooden board with geometric cutouts
- **Object B:** Empty
  Plain wooden surface
- **Object C:** Full
  Wooden board with filled shapes

Hold-out Object (D)

- **Object D:** Big cutout
  Larger wooden board with single large cutout
- Tested in WS4 only
- Never seen during training

Configuration Generalization Test

- **Train:** WS1, WS3 (Objects A,B,C)
- **Validate:** WS2 (Objects A,B,C)
- Tests **configuration generalization**
- Result: **65-75%** (partial)

Object Generalization Test

- **Train:** WS1,2,3 (Objects A,B,C)
- **Hold-out:** WS4 (Object D)
- Tests **object generalization**
- Result: **50%** (random)

Backup: The Entanglement Problem Explained

The Entanglement Problem

Why Acoustic Signatures Cannot Separate Contact from Object Properties

Two Entanglement Problems

1. **Signal = Contact ⊗ Configuration**
Robot pose affects acoustic response

2. **Signal = Contact ⊗ Object**
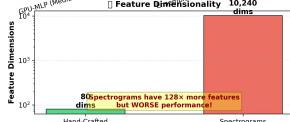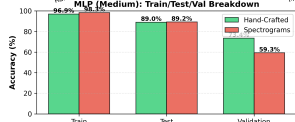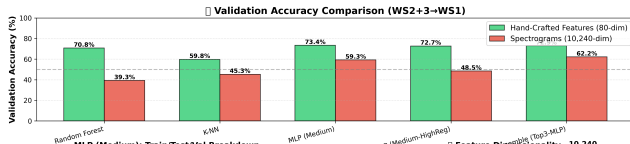Object identity entangled with contact

Physics Explanation

- Robot pose affects wave propagation
- Each object has unique eigenfrequencies
- Material, size, geometry affect response

Implications

- Not failures — **discoveries**

Backup: Hand-Crafted Features vs Spectrograms

Hand-Crafted Features vs Spectrograms: Complete Comparison

Key Results

Validation Accuracy (WS2+3→WS1):

- **Hand-Crafted: 73.4%**
- Spectrograms: 59.3%
- **+14.1% improvement!**

Why Hand-Crafted Wins

- **128× fewer features** (80 vs 10,240)
- **Less overfitting** with limited data
- **More interpretable** features
- **Faster inference** (¡1ms)
- **Better generalization**