

# LEARNING MULTIMODAL SIM-TO-REAL ROBOT POLICIES WITH GENERATIVE AUDIO

Tech ID: 34326 / UC Case 2026-053-0

## PATENT STATUS

Patent Pending

## BRIEF DESCRIPTION

The deployment of robotic systems in real-world environments is often limited by the "sim-to-real gap," where policies trained in digital simulations fail to account for the complex, multisensory feedback of physical reality. Researchers at UC Berkeley have developed a novel method for training multimodal sim-to-real robot policies by integrating generative audio models with traditional physics-based simulators. This framework uses a generative model to synthesize realistic audio data that corresponds to simulated physical interactions, creating a rich, multimodal dataset for policy learning. By training on both simulated physics and generated sensory data, the system enables robots to develop more robust and adaptive behaviors that translate seamlessly from virtual training environments to complex real-world tasks.

## SUGGESTED USES

»

Industrial Automation: Training assembly robots to detect hardware seating or mechanical alignment through acoustic feedback.

»

Service Robotics: Enhancing the ability of domestic robots to interact with fragile objects or varied floor surfaces using sound.

»

Autonomous Exploration: Providing robots in remote or low-visibility environments with supplemental sensory data for navigation.

»

Quality Control: Implementing acoustic monitoring for robotic sorting systems to identify material defects or structural inconsistencies.

»

Human-Robot Interaction: Improving robot responsiveness to environmental cues in collaborative workspaces.

## ADVANTAGES

»

Multimodal Robustness: Bridges the sim-to-real gap by providing robots with realistic sensory modalities beyond simple vision or touch.

»

Data Efficiency: Leverages generative models to create vast amounts of labeled sensory data without the need for expensive, time-consuming real-world collection.

»

Adaptive Learning: Enables policies to account for environmental noise and acoustic signatures that affect real-world performance.

## CONTACT

Laleh Shayesteh  
lalehs@berkeley.edu  
tel: 510-642-4537.



## INVENTORS

» Anumanchipalli, GopalaKrishna

## OTHER INFORMATION

### CATEGORIZED AS

» **Computer**

» Software

» **Engineering**

» Engineering

» Robotics and Automation

### RELATED CASES

2026-053-0

»

Scalable Framework: Compatible with standard physics-based simulators and a wide variety of generative sensory models.

»

Enhanced Perception: Allows robots to "hear" physical properties of objects—such as hollowness or material type—that are difficult to perceive through vision alone.

## RELATED MATERIALS

---

### ADDITIONAL TECHNOLOGIES BY THESE INVENTORS

- ▶ [Realtime Transformation Of Voice For Privacy Protection](#)
- ▶ [Realtime Transformation Of Voice Identity And Style](#)
- ▶ [Methods For Dysfluent Speech Transcription And Detection](#)
- ▶ [Articulatory Feedback For Phonetic Error-Based Pronunciation Training](#)



University of California, Berkeley Office of Technology Licensing

2150 Shattuck Avenue, Suite 510, Berkeley, CA 94704

Tel: 510.643.7201 | Fax: 510.642.4566

<https://ipira.berkeley.edu/> | [otl-feedback@lists.berkeley.edu](mailto:otl-feedback@lists.berkeley.edu)

© 2026, The Regents of the University of California

[Terms of use](#) | [Privacy Notice](#)