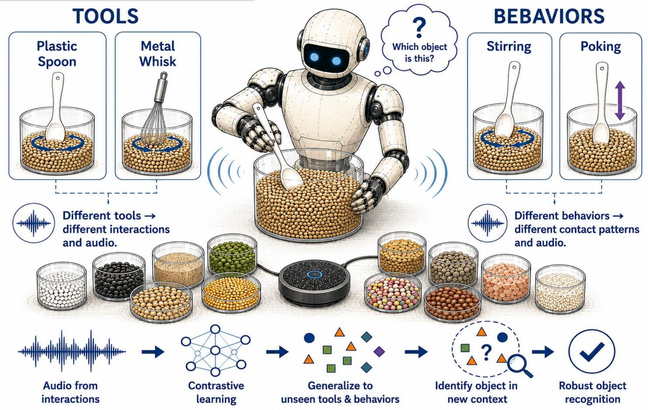
Contrastive Auditory Knowledge Transfer for Tool-Mediated Robot Interaction with Granular Objects

Si Liu[1], Jindan Huang[1], Zhengyan Huan[2], Michael C. Hughes[1], Jivko Sinapov[1]

[1]Department of Computer Science; [2]Department of Electrical and Computer Engineering, School of Engineering, Tufts University, Medford, MA,USA. {firstname.lastname}@tufts.edu

1. Motivation

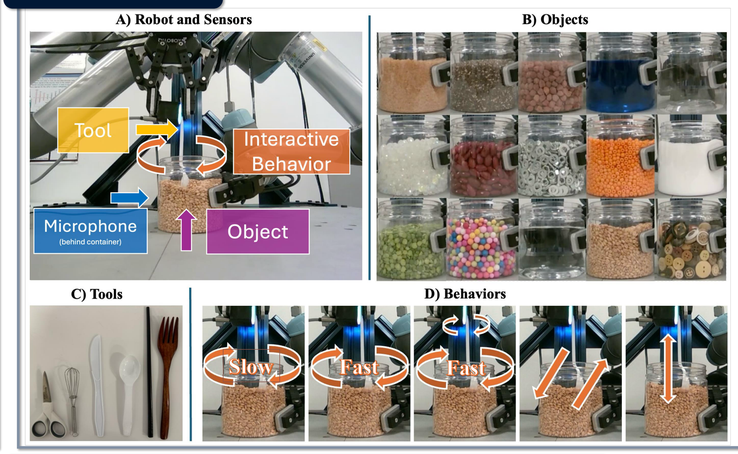


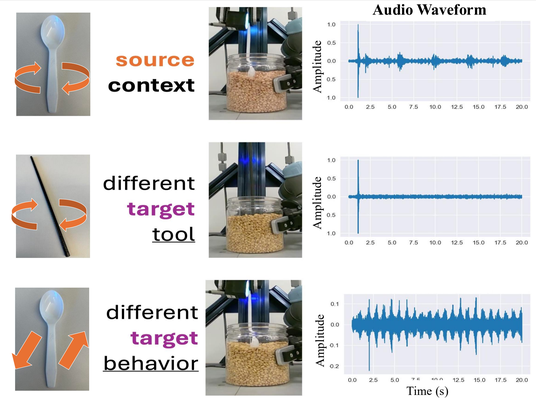
Tool-mediated interactions enable robots to manipulate and explore granular objects, producing informative auditory signals.

A central challenge is transferring this perceptual knowledge across different tools and behaviors without costly data collection for each new context.

In this work, we leverage audio signals from tool-mediated interactions and use contrastive learning for knowledge transfer.

2. Dataset





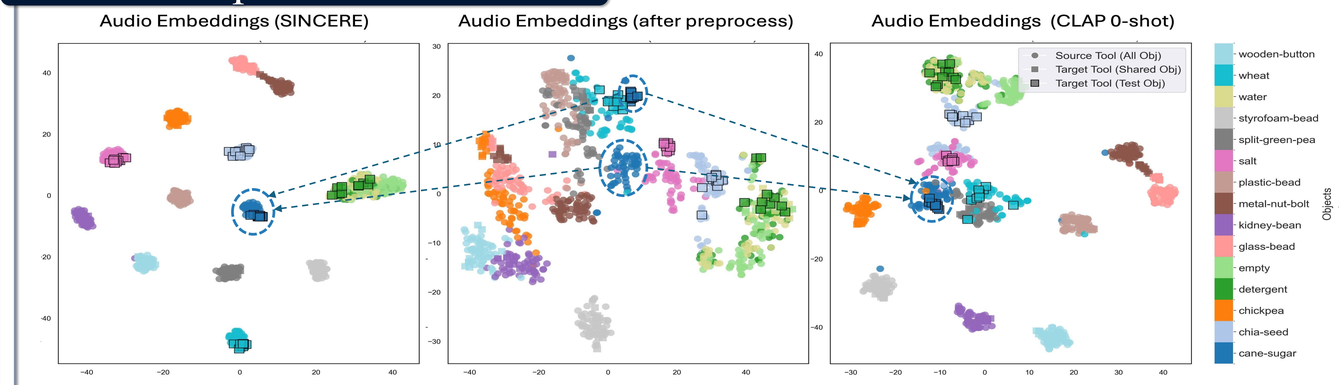
Models trained in **source** contexts often fail in a **target** context, due to domain shift.

3. Knowledge Transfer Overview

Left: The model learns object-level features invariant to the context, so target-tool interactions can be mapped to the same object clusters learned from the source tool.

Right: Novel objects can be inferred by aligning sensory embeddings with interaction descriptions.

4. Feature Space Visualization

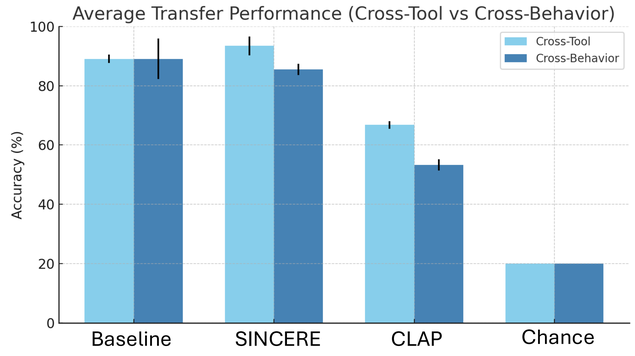


Starting from pretrained audio features (middle), our methods (left & right) better align transfer-object embeddings with their object clusters, improving target-context generalization. Even though the CLAP encoder hasn't seen the transfer objects during training, it still projects most test samples close to their corresponding clusters (right).

5. Transfer Pipeline

Left (SINCERE): source and target embeddings are projected into a shared latent space and clustered by object.
Right (CLAP): audio embeddings are aligned with text embeddings, enabling prediction by nearest text similarity.

6. Result



SINCERE shows effective transfer when source contexts include test objects. CLAP enables above-chance 0-shot recognition of novel objects.

7. Future Work

- Add multimodal sensing with tactile and force feedback
- Use prompt learning to improve text-guided transfer
- Test online on real robots for real-time inference
- Collect broader datasets across robots and environments

MuLIP Multimodal Learning, Interaction, and Perception Lab

ICRA IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION 2026 VIENNA