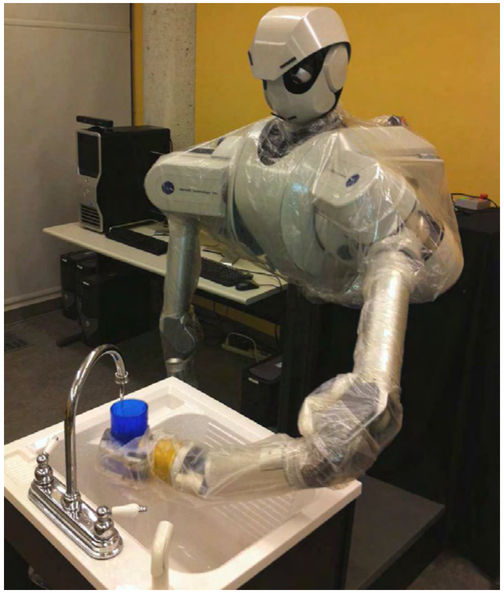
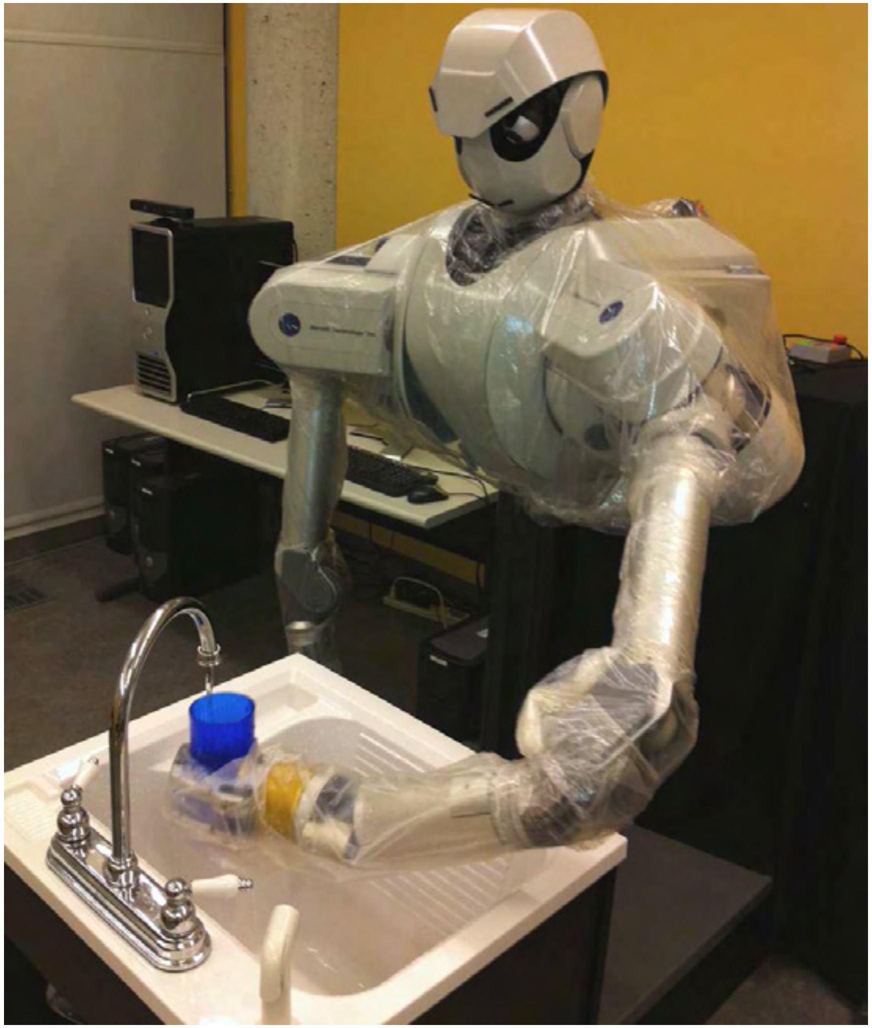
Object Categorization in the Sink:
Learning Behavior–Grounded Object Categories with Water



Shane Griffith, Vlad Sukhoy, Todd Wegter, and Alex Stoytchev
Developmental Robotics Laboratory
Iowa State University
www.ece.iastate.edu/~shaneg

Humanoid Robots and Water *Can* Play Well Together



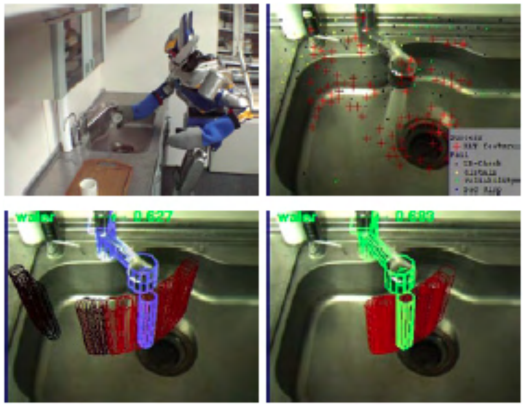
Water Use is Universal

Cooking **Cleaning** **Gardening**

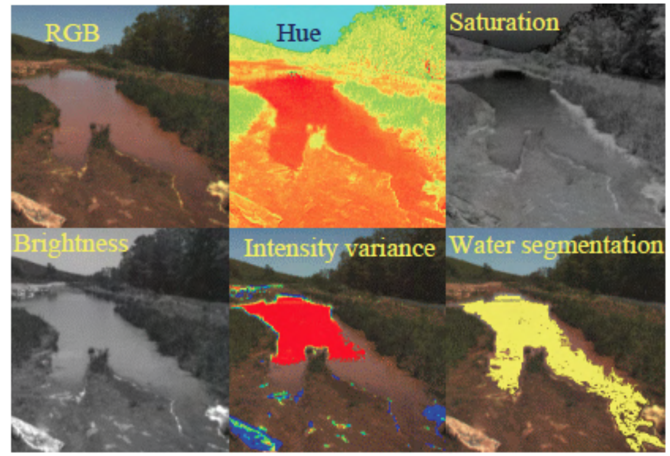




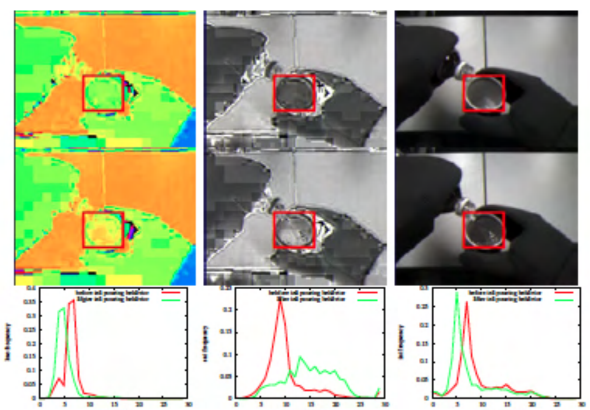

How to Observe Water using Vision?



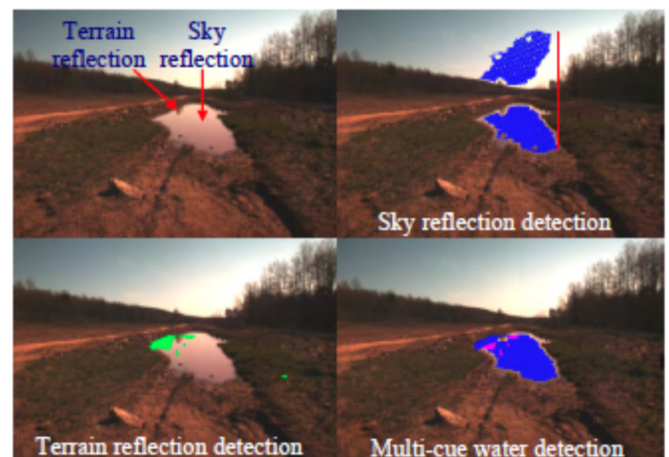
Okada *et al.*; 2009



Rankin and Matthies; 2010



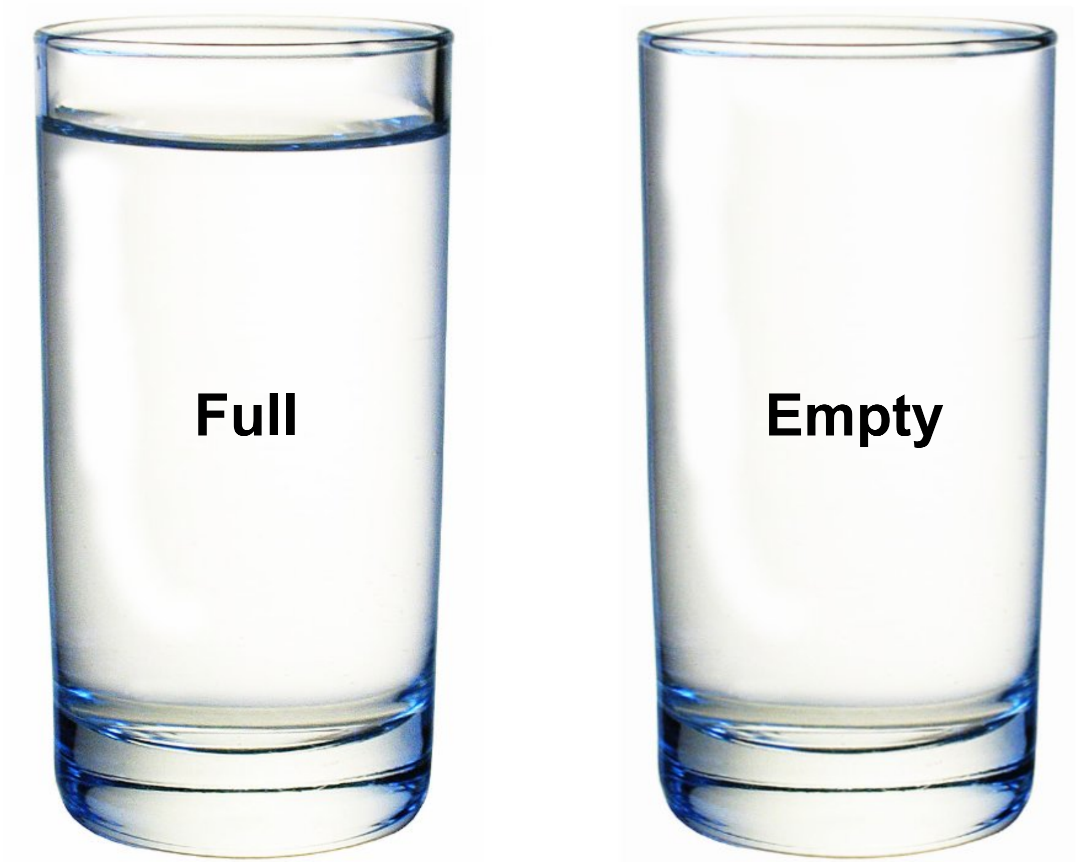
Okada *et al.*; 2009



Rankin, Matthies, and Bellutta; 2011

Water is Hard to See

Water is Hard to See



Some Objects That Hold Water Just Aren't Easy to Identify

Object Shape Can Be Deceptive






Research Question

What is a container?

Infant Interacting with a Container

Infant Interacting with a Non-Container

Previous Work

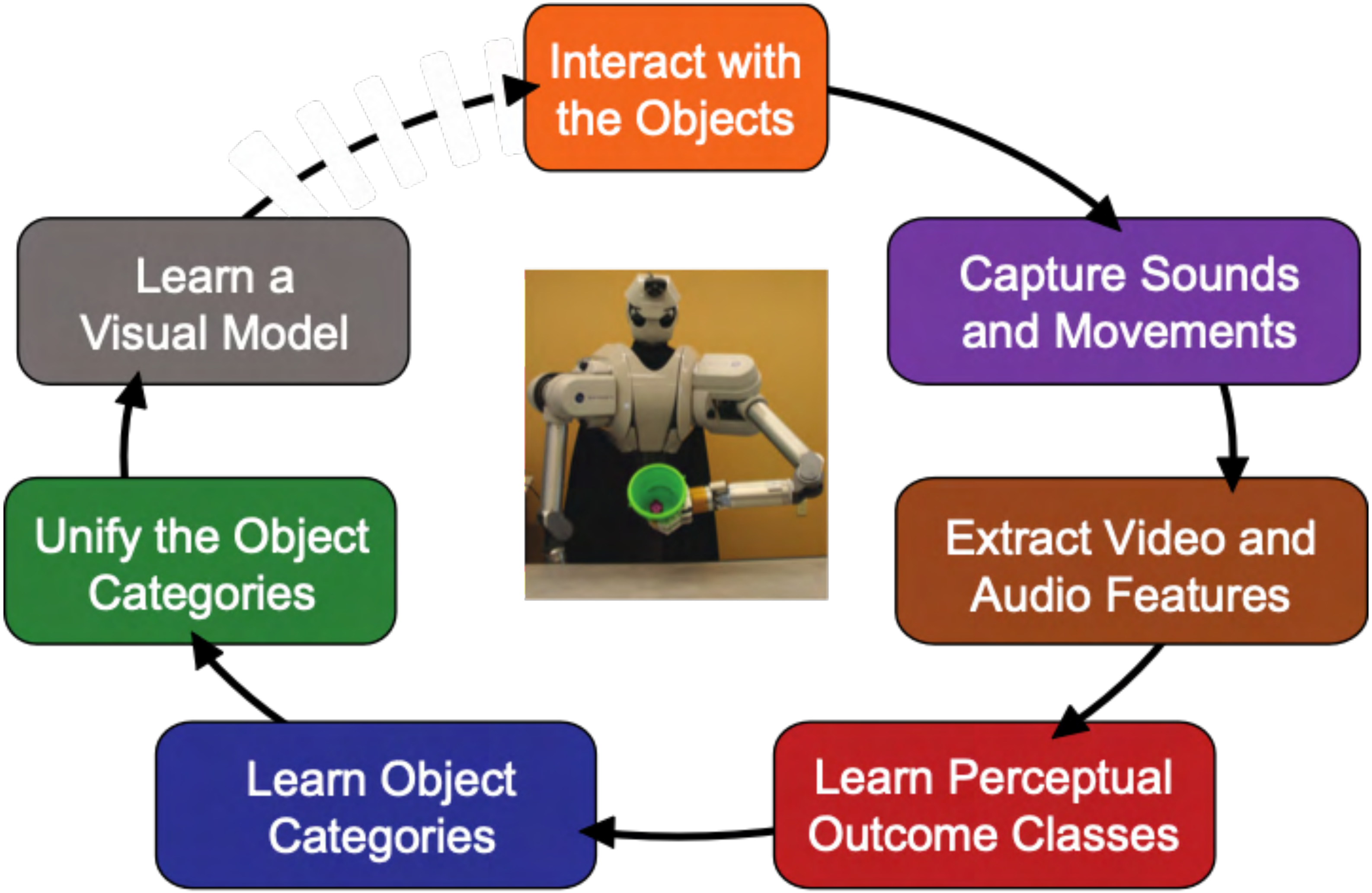
Vision and Audio

Griffith *et al.*; 2012

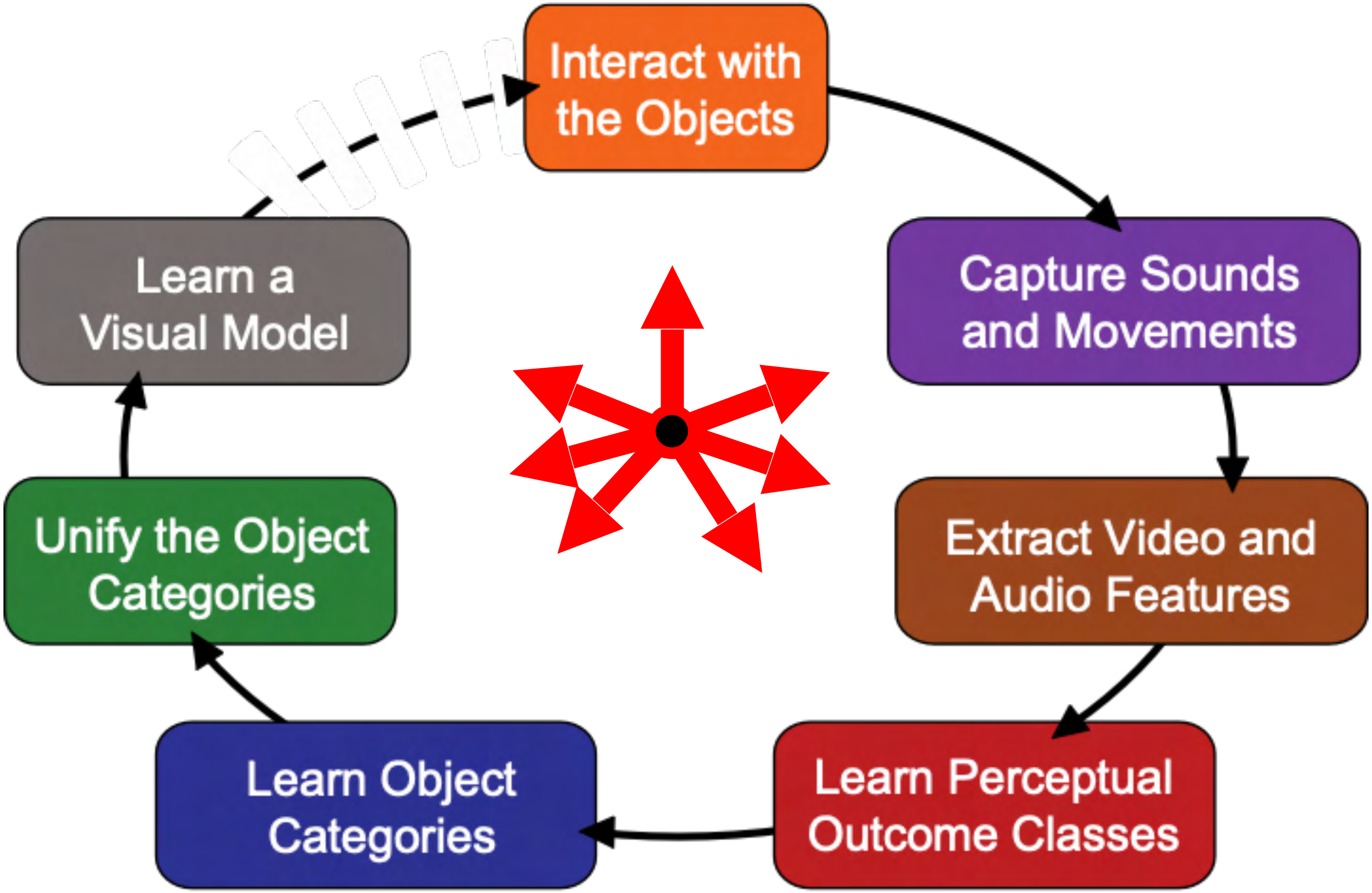
Audio and Proprioception

Sinapov *et al.*; 2011

Learning Framework

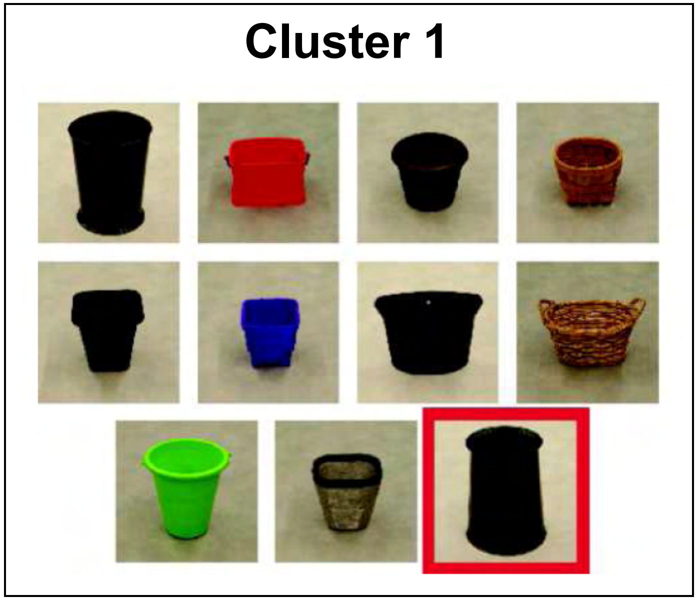


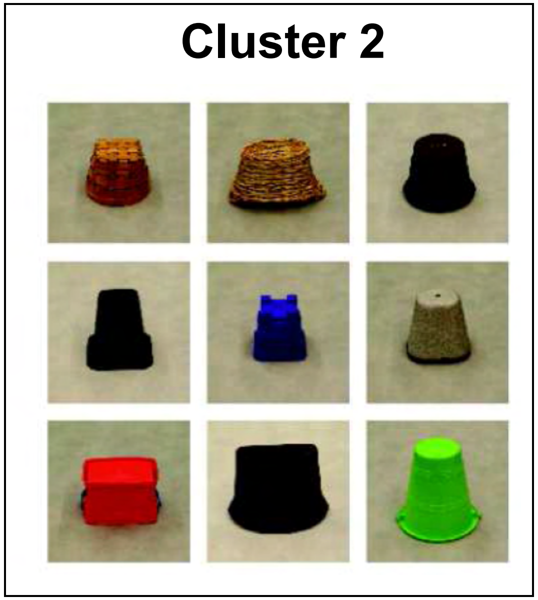
Learning Framework



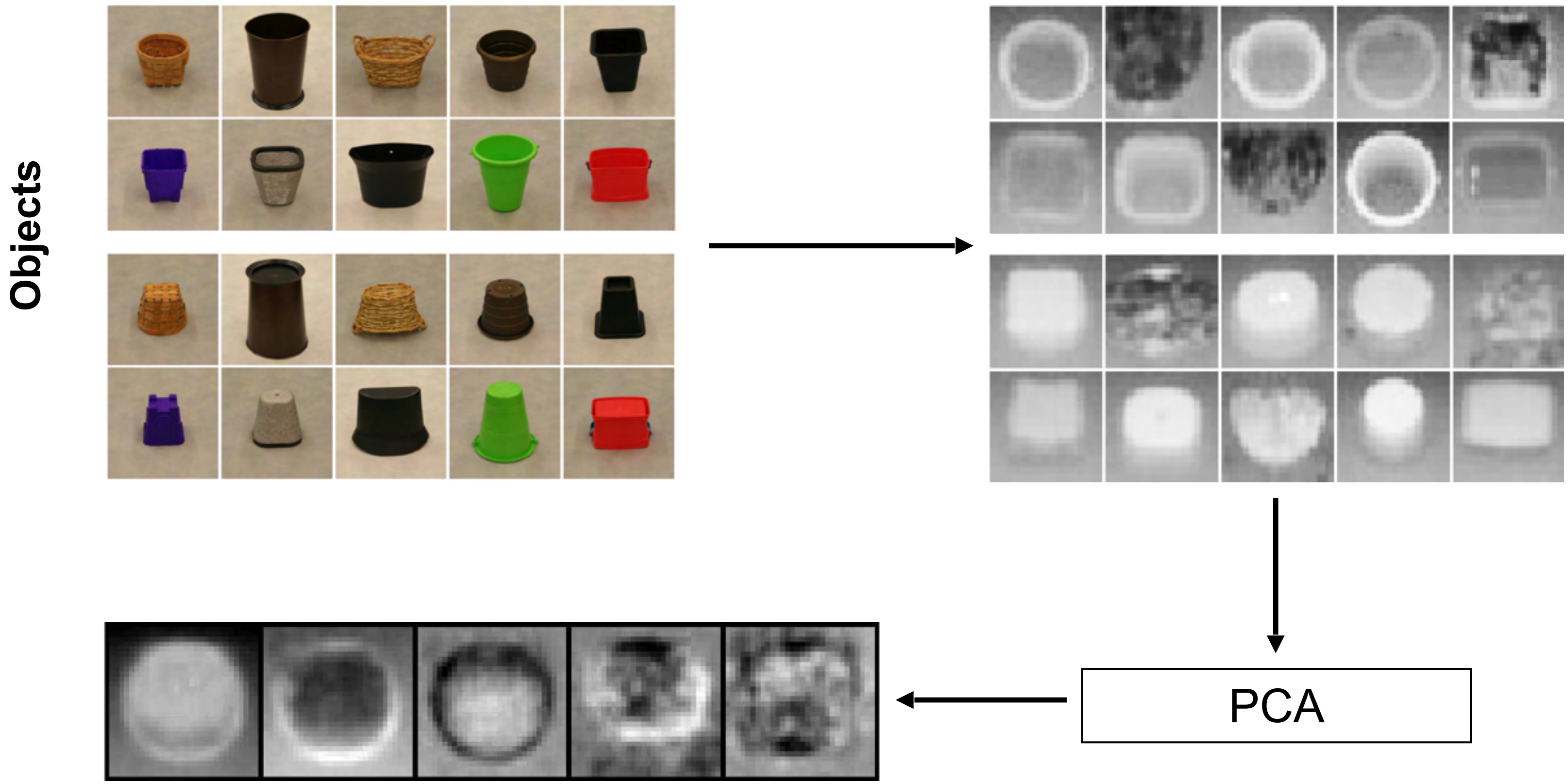
Unified Categorization

(derived from both sound and movement observations)



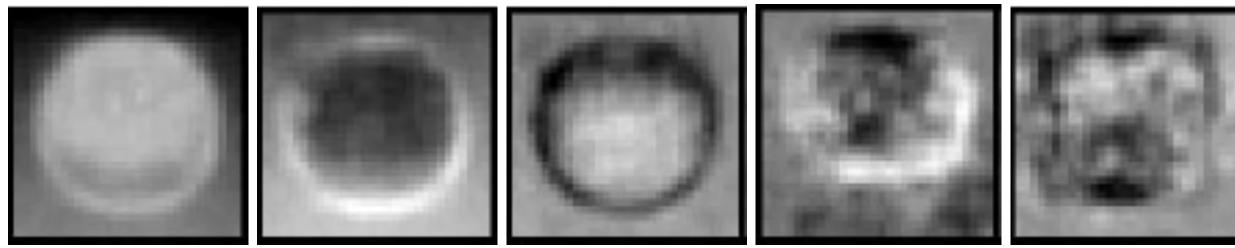


Extracted Visual Features



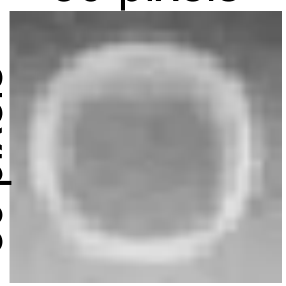
learned features

Example Visual Feature Activation



30 pixels

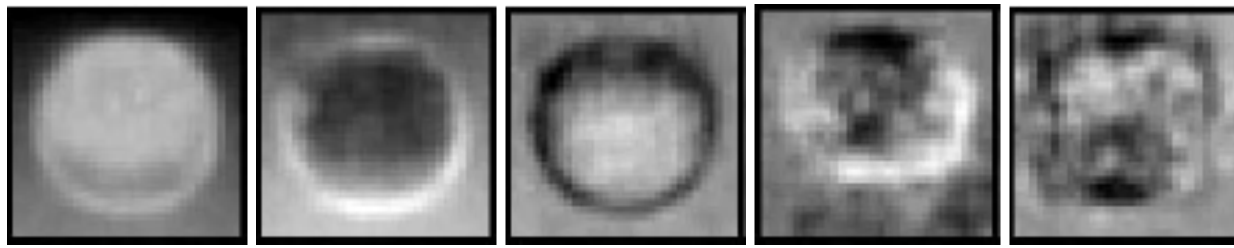
30 pixels



.19 .28 -.3 .02 .12

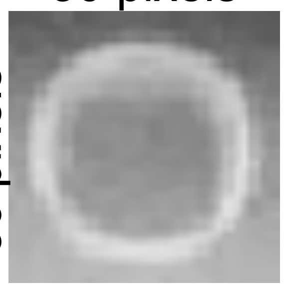
900 values → **PCA** → **5 values**

Example Container Feature Activation



30 pixels

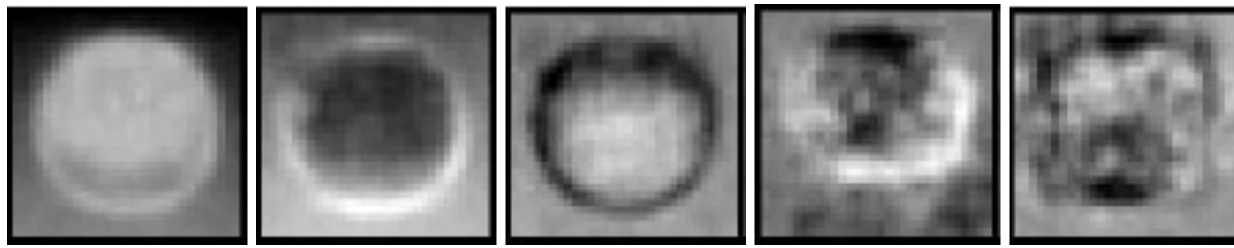
30 pixels



.19	.28	-.3	.02	.12
Strongly activated	Strongly activated	Strongly activated	Weakly activated	Weakly activated

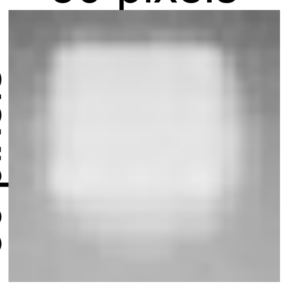
The concave features are the most strongly activated

Example Non-Container Feature Activation



30 pixels

30 pixels



.31	-.14	.1	.04	-.08
Strongly activated	Weakly activated	Weakly activated	Weakly activated	Weakly activated

The convex feature is the most strongly activated

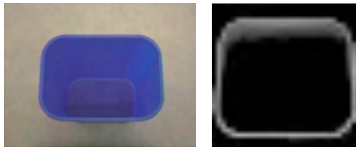
Classifying Novel Objects

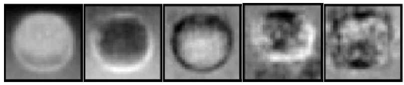
Training Objects

					class label
0.19	0.28	-0.3	0.02	0.12	C_1
0.02	0.29	0.17	-0.03	0.67	C_1
0.22	0.2	-0.36	0.12	-0.06	C_1
0.2	0.32	0.0	0.04	0.0	C_1
-0.02	0.31	0.29	0.1	-0.22	C_1
0.18	0.31	0.0	-0.18	-0.13	C_1
0.28	0.13	-0.29	0.03	-0.03	C_1
0.02	0.27	0.41	-0.15	0.29	C_1
0.22	0.18	-0.4	0.0	0.07	C_1
0.04	0.35	0.09	0.04	-0.18	C_1

					class label
0.31	-0.14	0.1	0.04	-0.08	C_2
0.12	0.25	0.29	-0.29	-0.52	C_1
0.31	-0.14	0.09	0.04	0.07	C_2
0.3	-0.18	0.09	-0.09	-0.03	C_2
0.22	0.04	0.3	0.56	0.12	C_2
0.29	-0.14	0.11	-0.26	0.11	C_2
0.29	-0.18	0.14	0.0	0.08	C_2
0.25	-0.15	0.16	0.44	-0.14	C_2
0.24	-0.17	0.04	-0.48	0.09	C_2
0.29	-0.14	-0.02	-0.08	-0.02	C_2

Novel Object





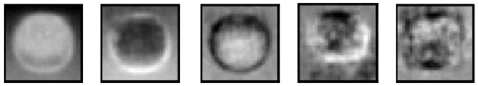
0.0 0.07 -0.2 -0.3 0.1

kNN

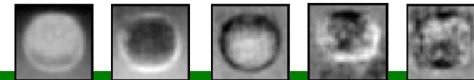
Cluster 1

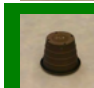
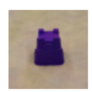
Classifying Novel Objects

Training Objects

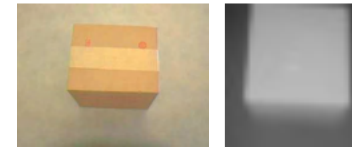


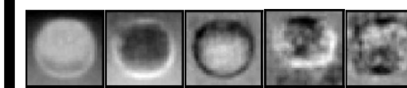
						class label
	0.19	0.28	-0.3	0.02	0.12	C_1
	0.02	0.29	0.17	-0.03	0.67	C_1
	0.22	0.2	-0.36	0.12	-0.06	C_1
	0.2	0.32	0.0	0.04	0.0	C_1
	-0.02	0.31	0.29	0.1	-0.22	C_1
	0.18	0.31	0.0	-0.18	-0.13	C_1
	0.28	0.13	-0.29	0.03	-0.03	C_1
	0.02	0.27	0.41	-0.15	0.29	C_1
	0.22	0.18	-0.4	0.0	0.07	C_1
	0.04	0.35	0.09	0.04	-0.18	C_1



						class label
	0.31	-0.14	0.1	0.04	-0.08	C_2
	0.12	0.25	0.29	-0.29	-0.52	C_1
	0.31	-0.14	0.09	0.04	0.07	C_2
	0.3	-0.18	0.09	-0.09	-0.03	C_2
	0.22	0.04	0.3	0.56	0.12	C_2
	0.29	-0.14	0.11	-0.26	0.11	C_2
	0.29	-0.18	0.14	0.0	0.08	C_2
	0.25	-0.15	0.16	0.44	-0.14	C_2
	0.24	-0.17	0.04	-0.48	0.09	C_2
	0.29	-0.14	-0.02	-0.08	-0.02	C_2

Novel Object





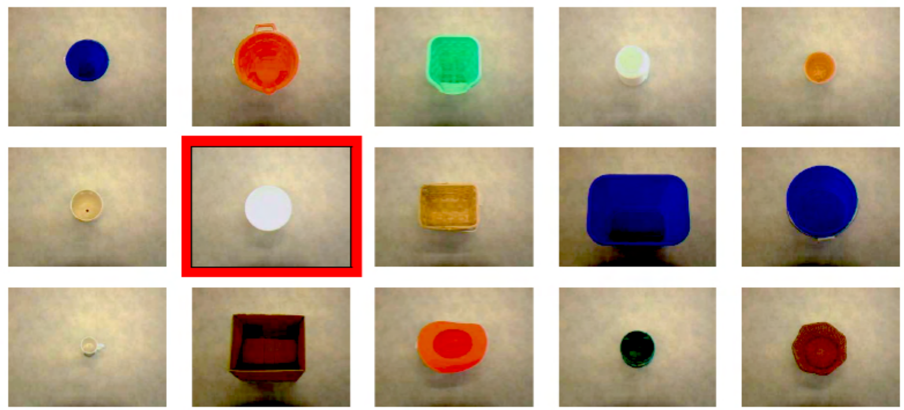
0.2 -0.2 0.0 0.05 -0.2

kNN

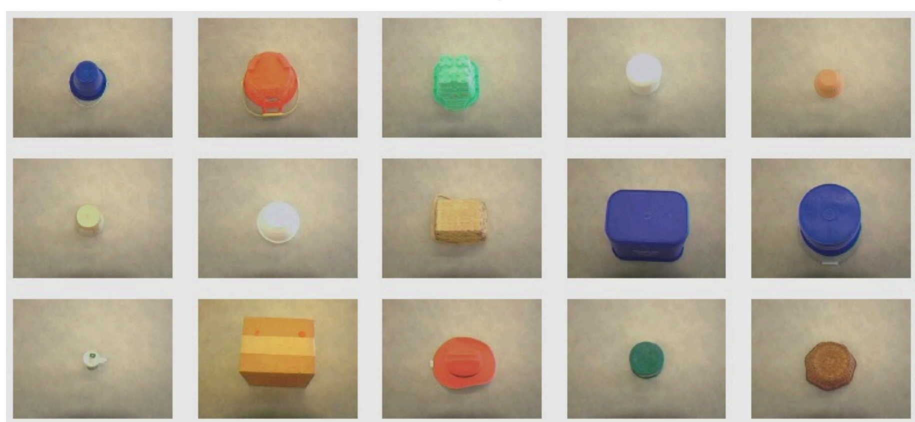
Cluster 2

Classification Results

Novel Containers



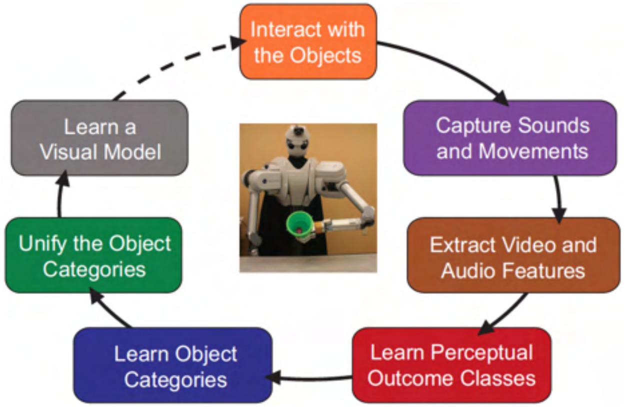
Novel Non-Containers



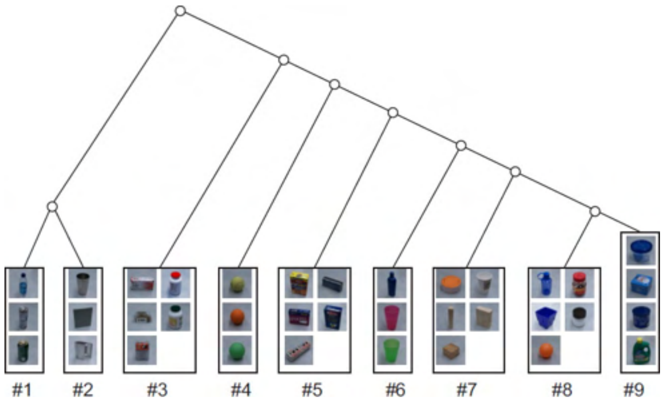
Infant Playing in the Sink

Video of the Experiments

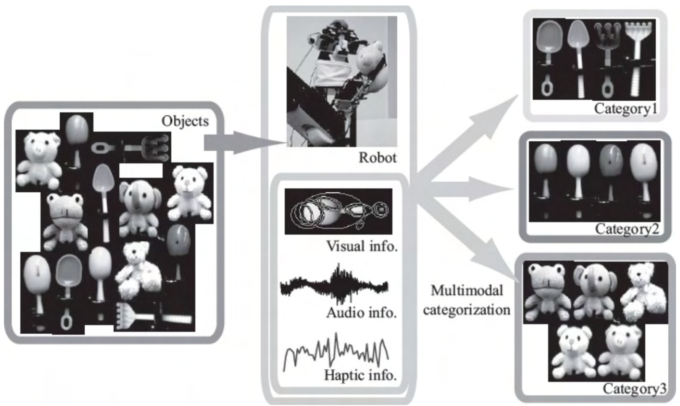
Interactive Object Categorization



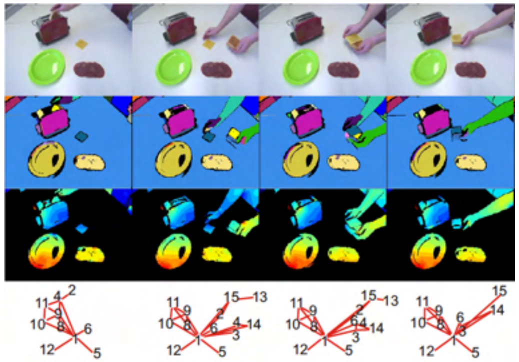
Griffith, Sinapov, Sukhoy, and Stoytchev; 2012



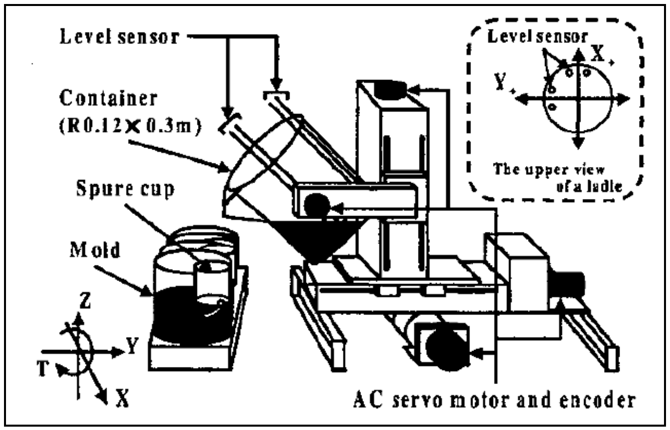
Sinapov and Stoytchev; 2009



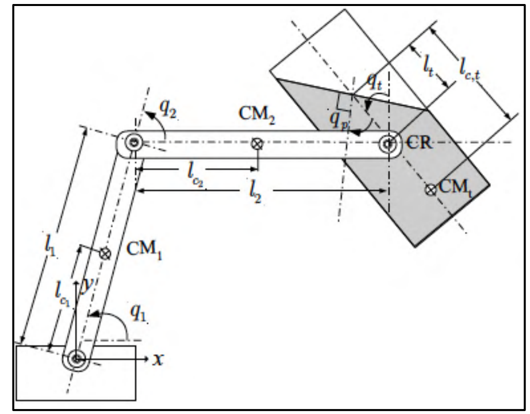
Nakamura, Nagai, and Iwahashi; 2007


Aksoy *et al.*; 2010

Slosh-Free Control of Containers

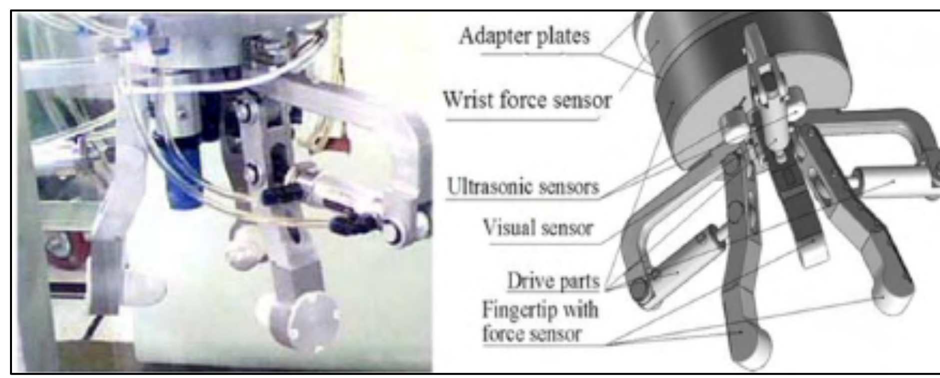


Yano *et al.*; 2001



Tzamtzi, Koumboulis, Kouvakas; 1997


Feddema *et al.*; 1997



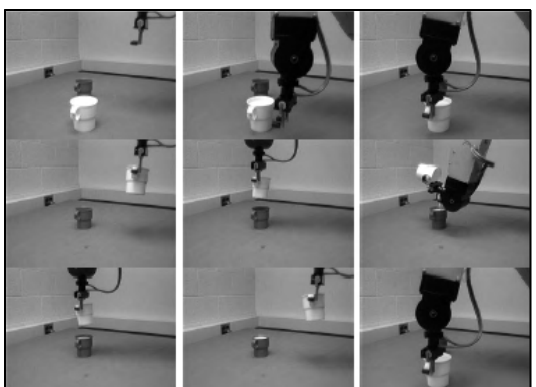
Liang, Zhang, Song, and Ge; 2010

Pouring Liquid into a Container

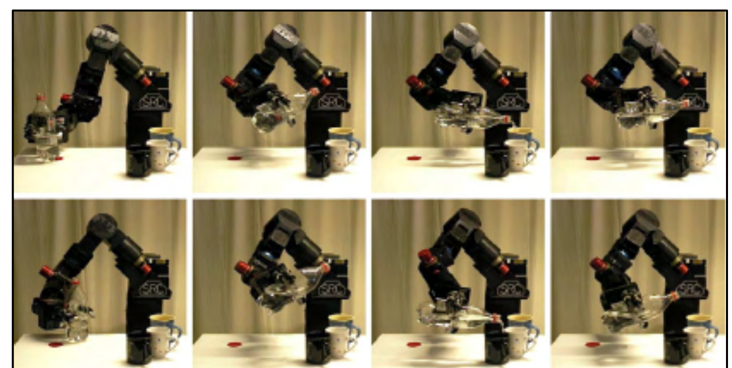
Cakmak and Thomaz; 2011

Okada *et al.*; 2009

Kim *et al.*; 2009

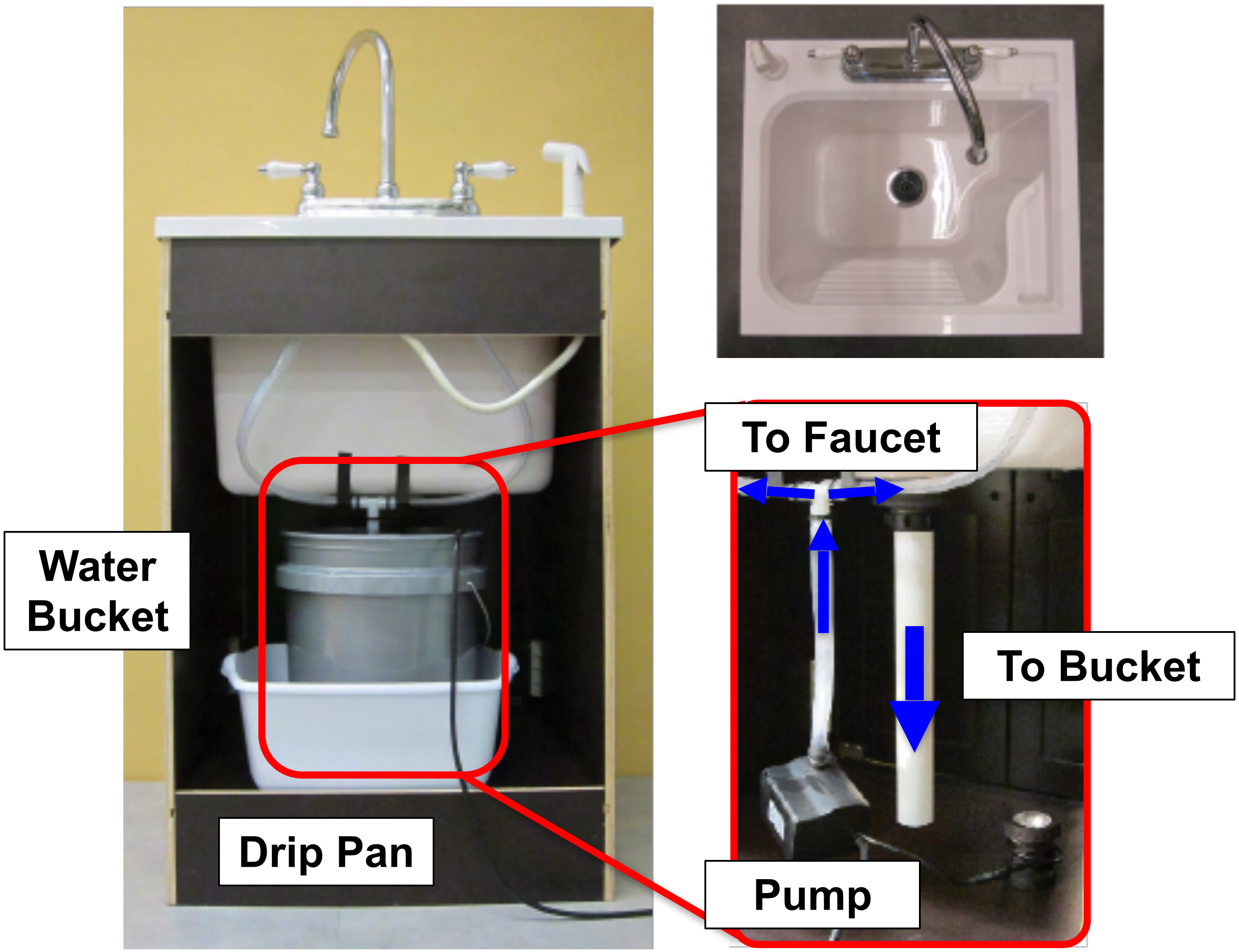


Hwang and Weng; 1997

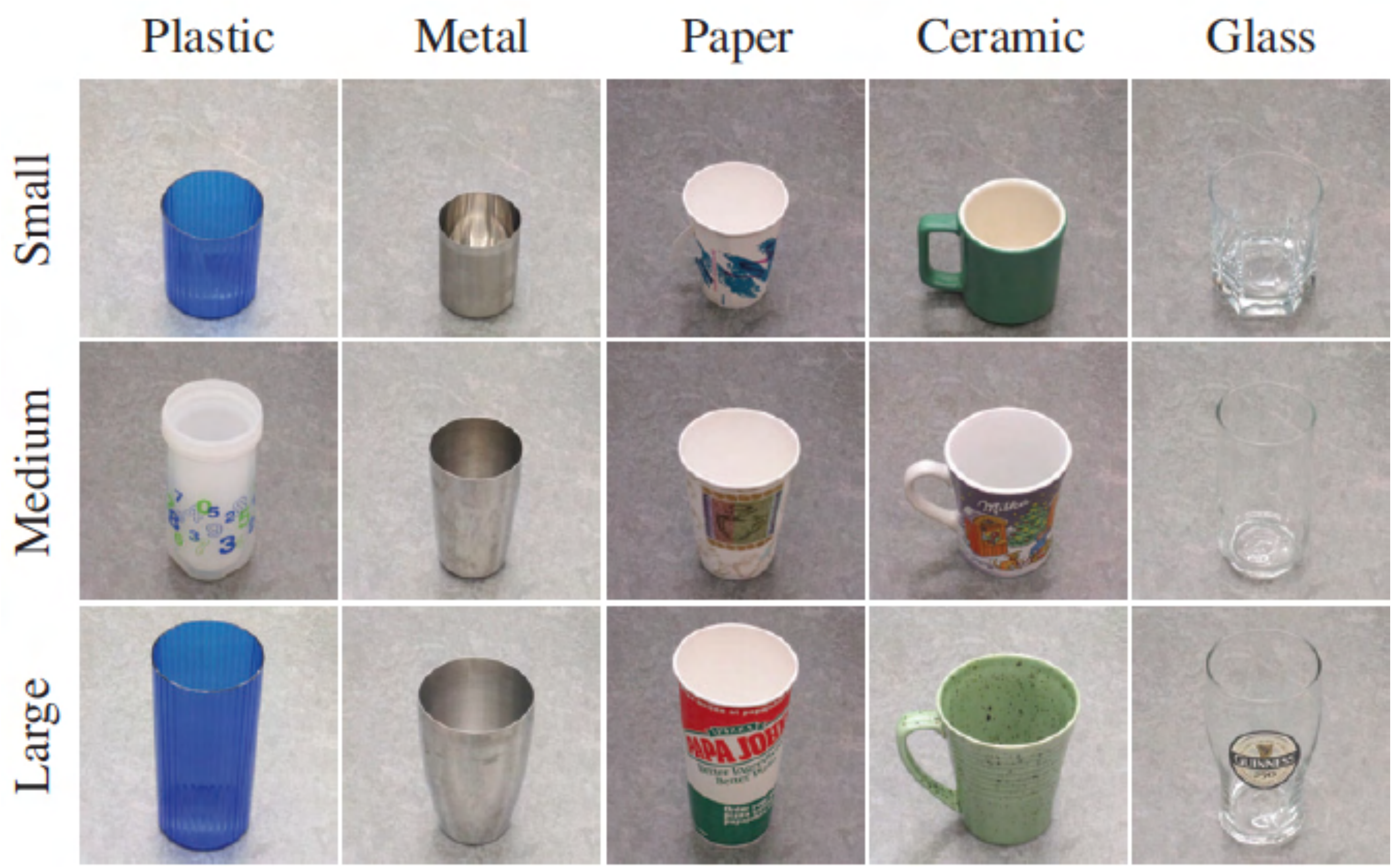


Pastor *et al.*; 2009

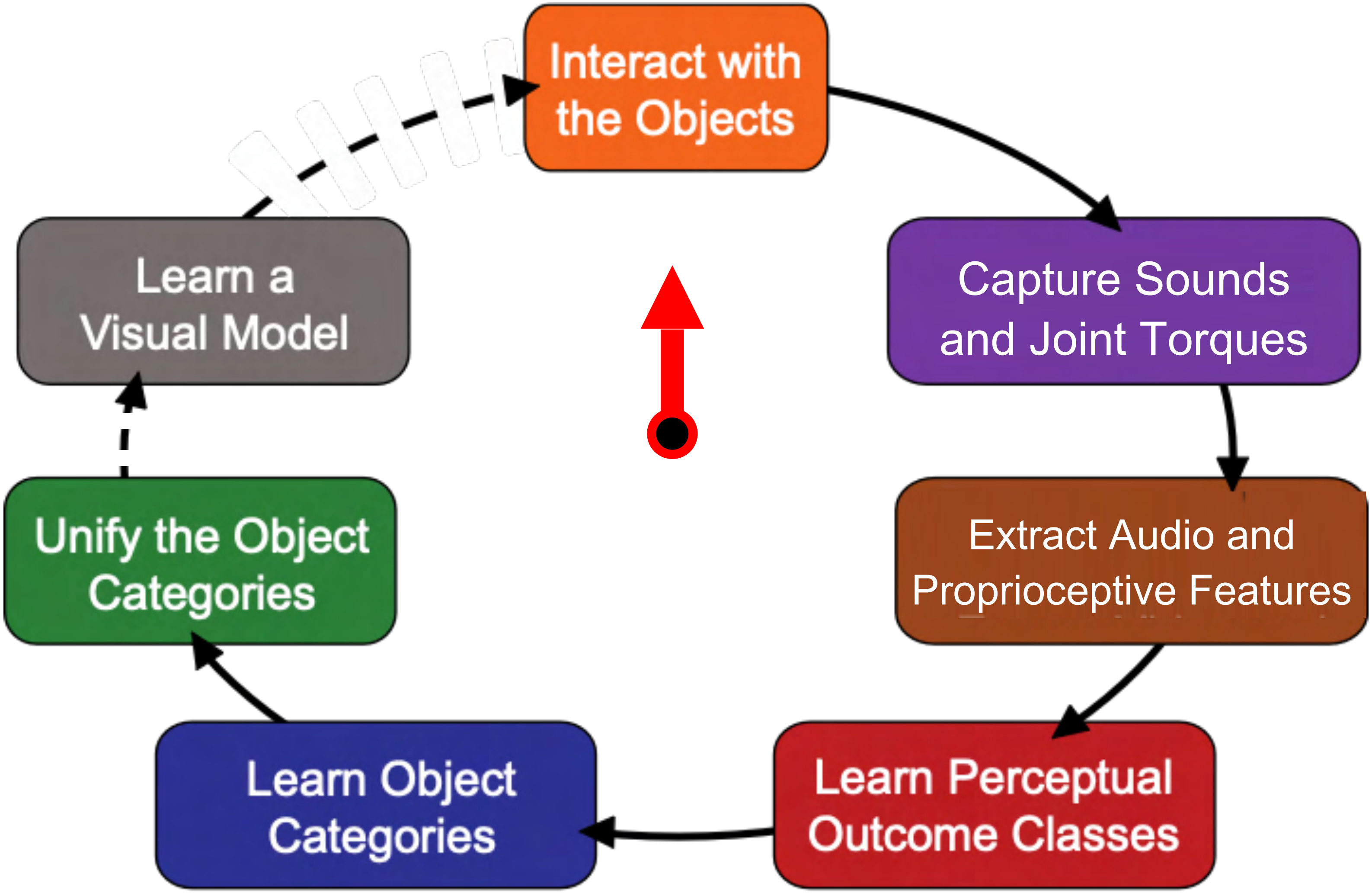
Sink



Can a Robot Categorize These Objects Using Audio and Proprioception?



Learning Framework



Behaviors

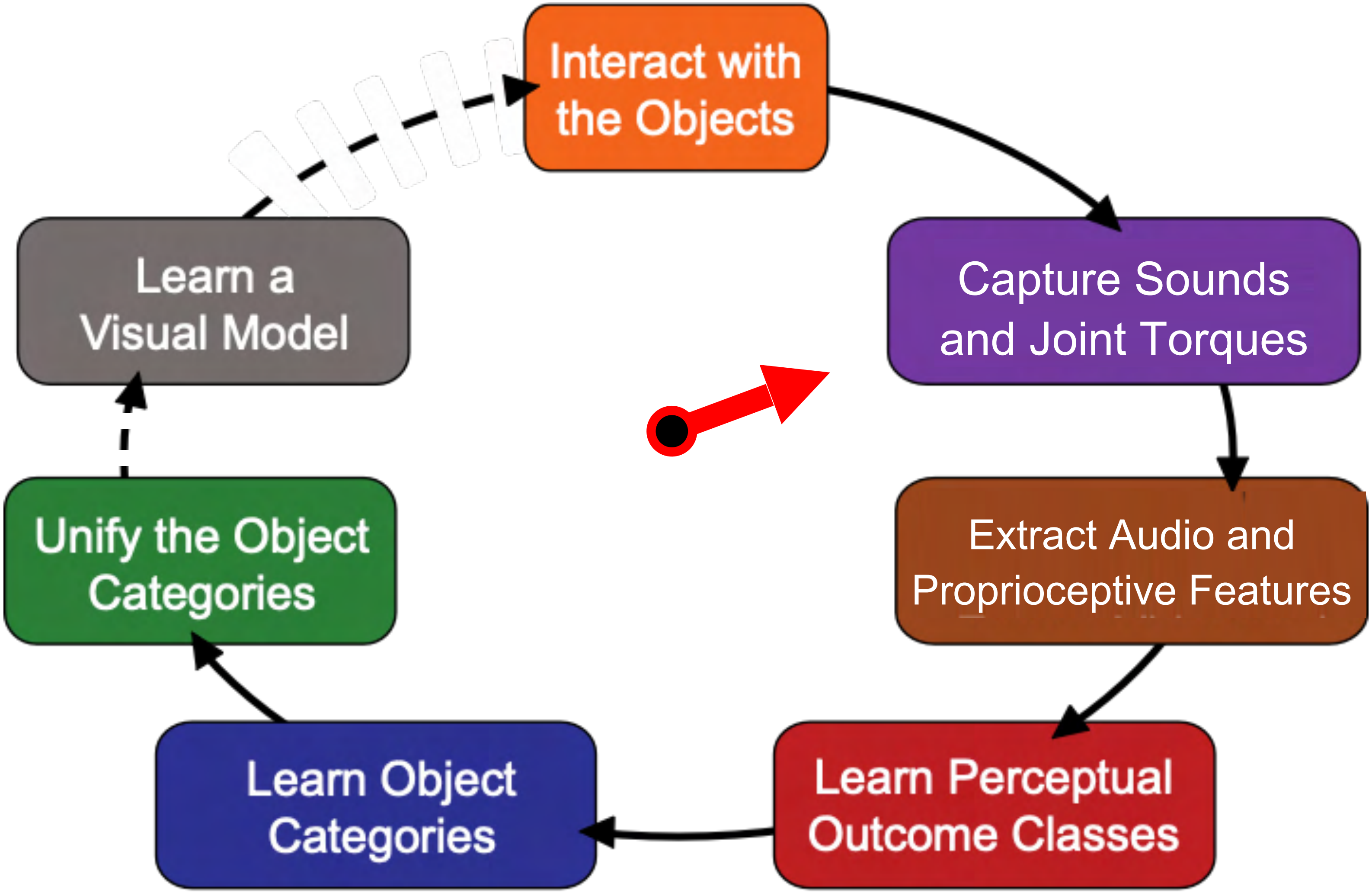
before
after

hold flip up and down rotate in and out

Don't Fry This At Home

Waterguard Cast and Skin Protector

Learning Framework



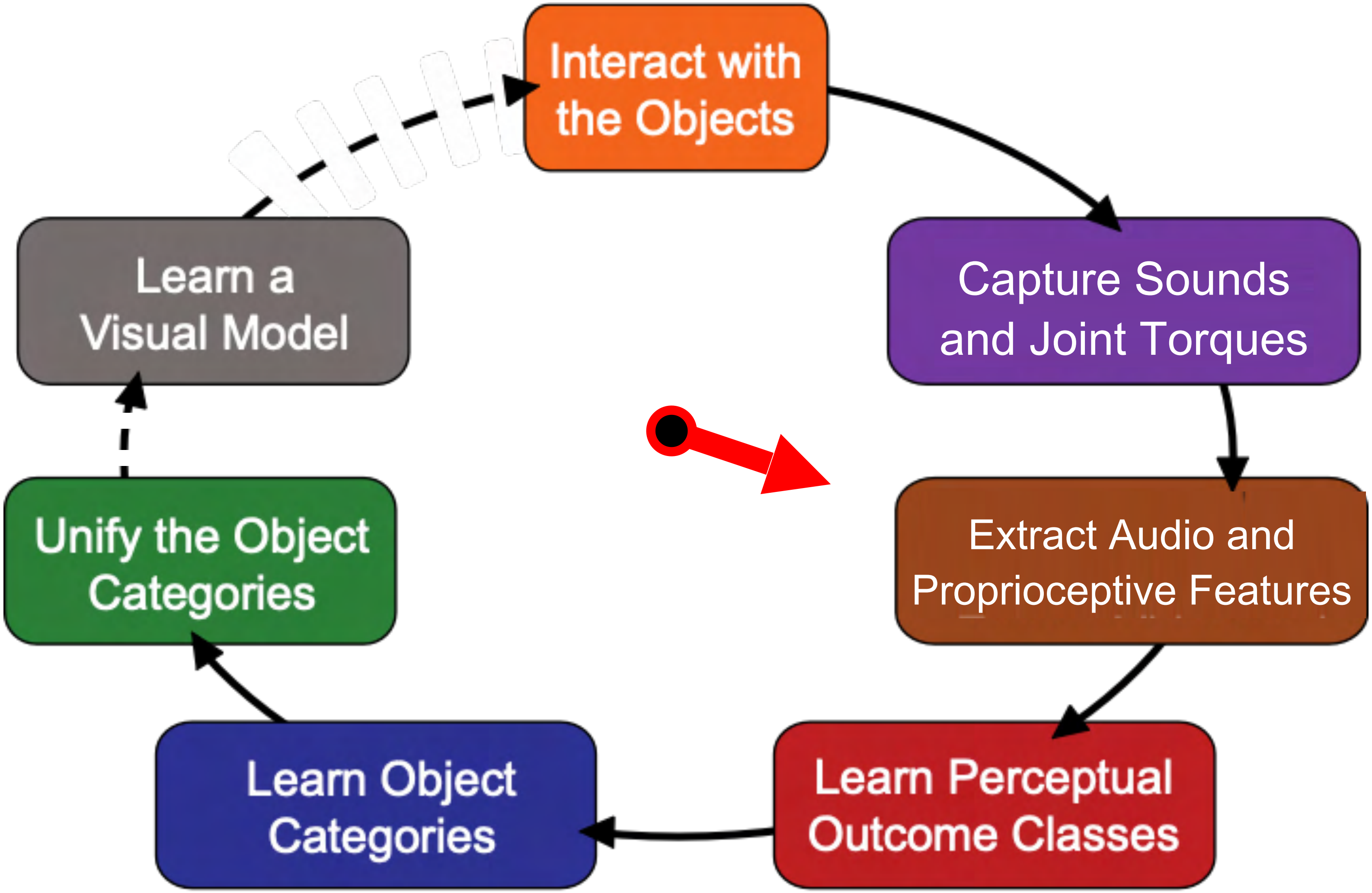
Data Collection

5 behaviors x 10 trials x 15 objects x 2 object poses

1,500 behavioral interactions

(6 hours of interaction)

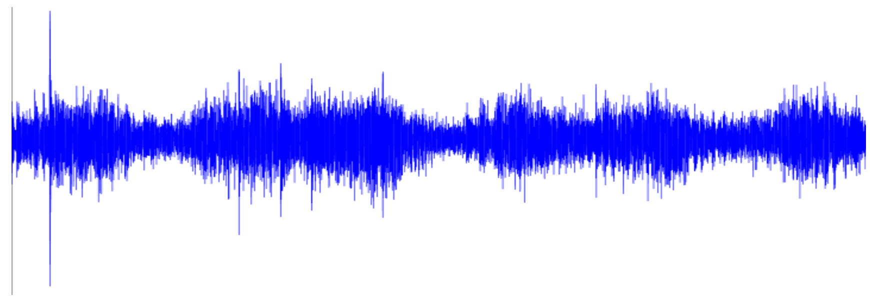
Learning Framework



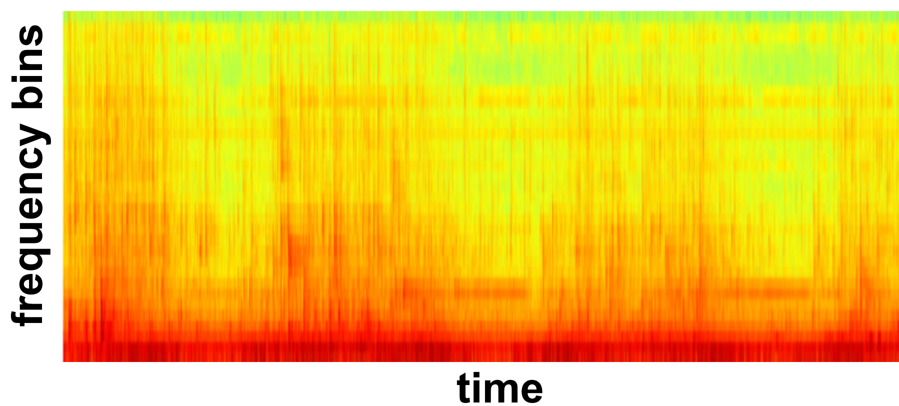
Audio Preprocessing

Behavior Execution:
(up and down)

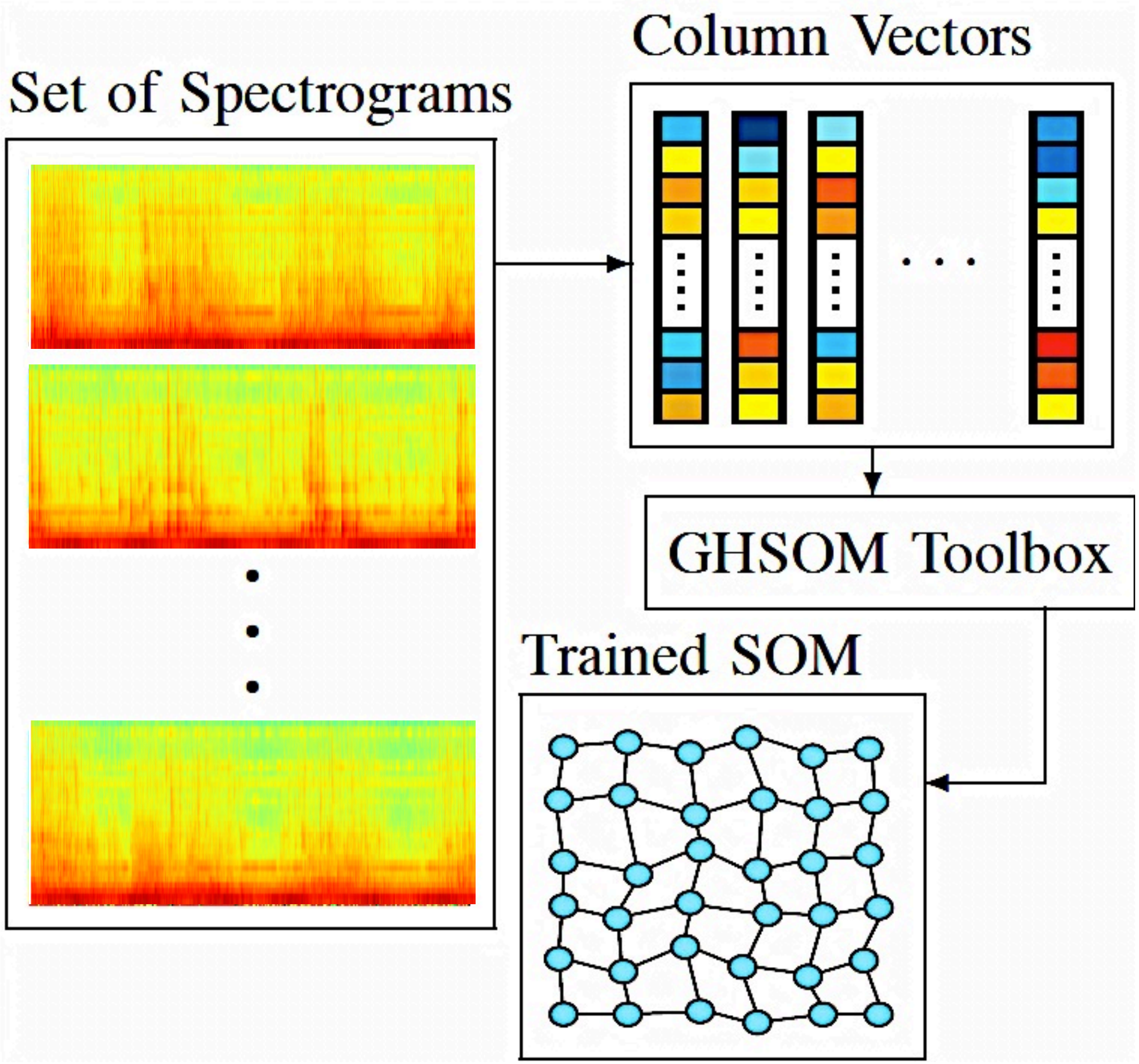
WAV file recorded:



Discrete Fourier Transform:

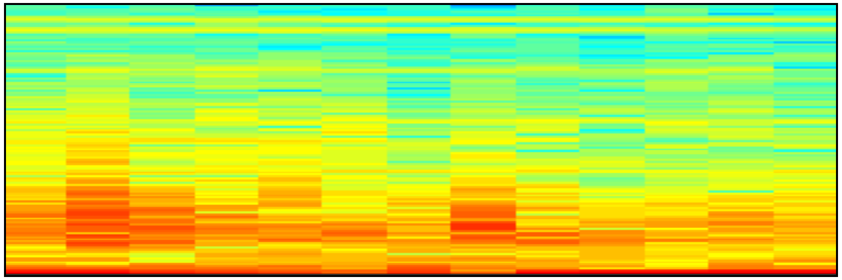


Unsupervised Feature Extraction

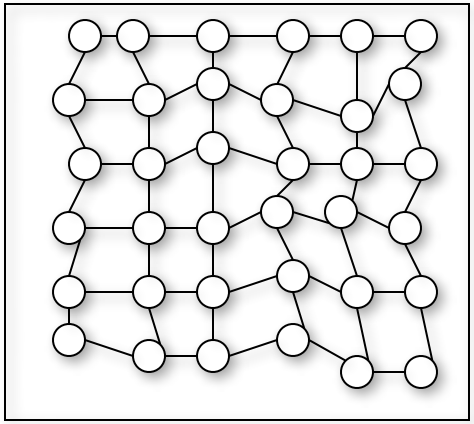


Convert the Spectrogram to a State Sequence Using a SOM

Spectrogram:



Self Organizing Map:

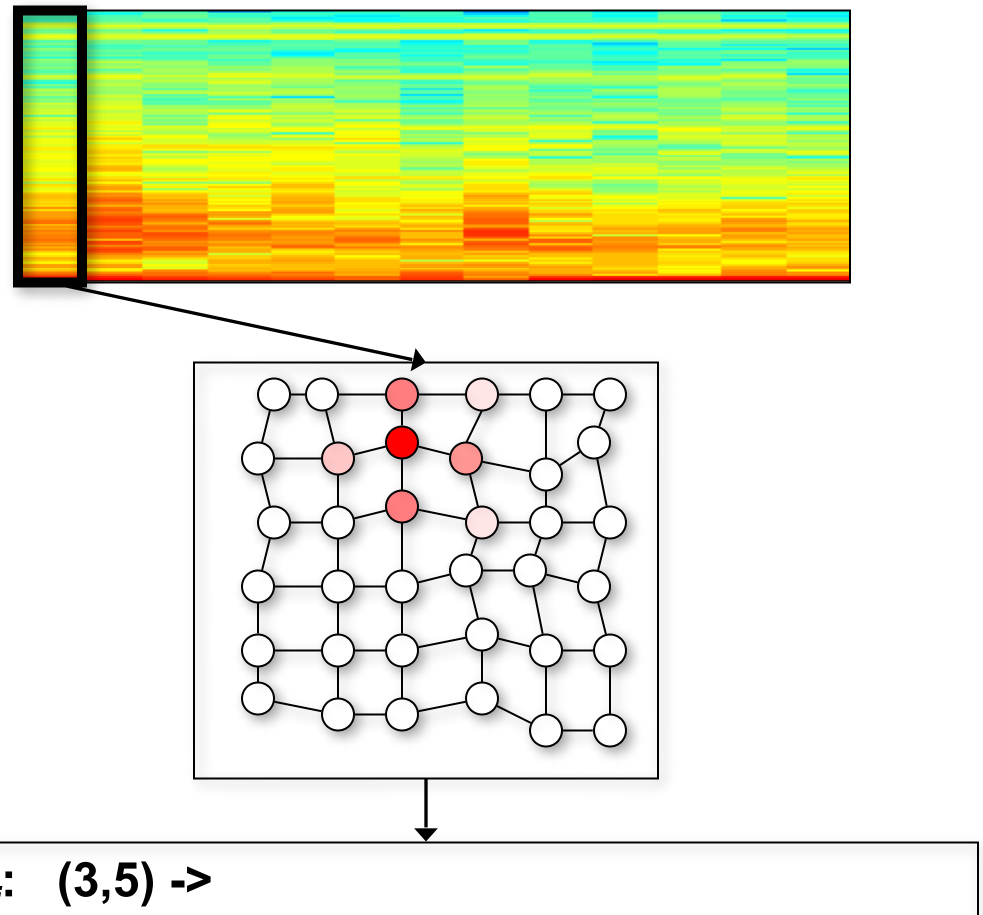


Convert the Spectrogram to a State Sequence Using a SOM

Spectrogram:

Self Organizing Map:

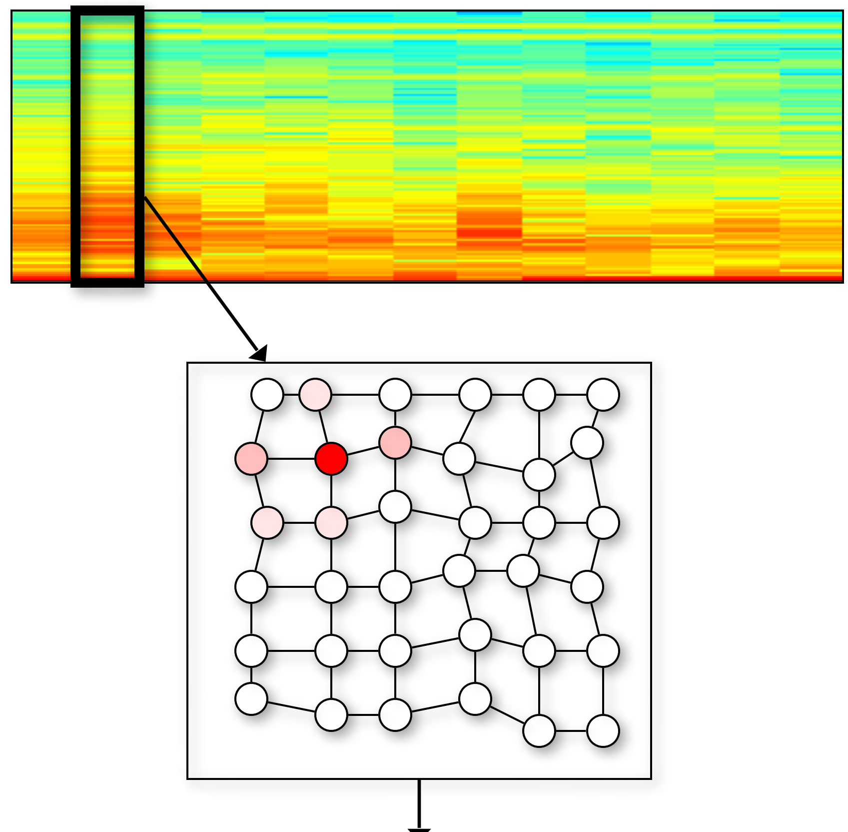
State Sequence:



Convert the Spectrogram to a State Sequence Using a SOM

Spectrogram:

Self Organizing Map:

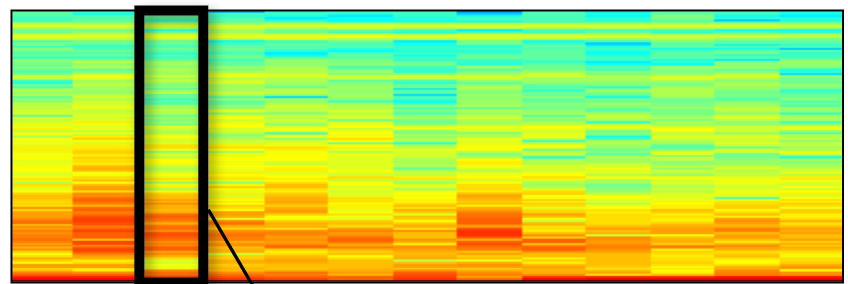


State Sequence:

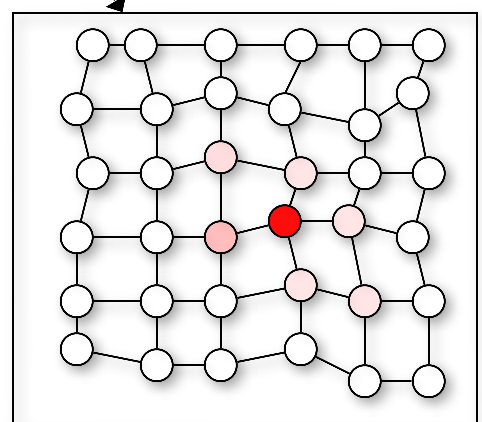
A_i: (3,5) -> (2,5) ->

Convert the Spectrogram to a State Sequence Using a SOM

Spectrogram:



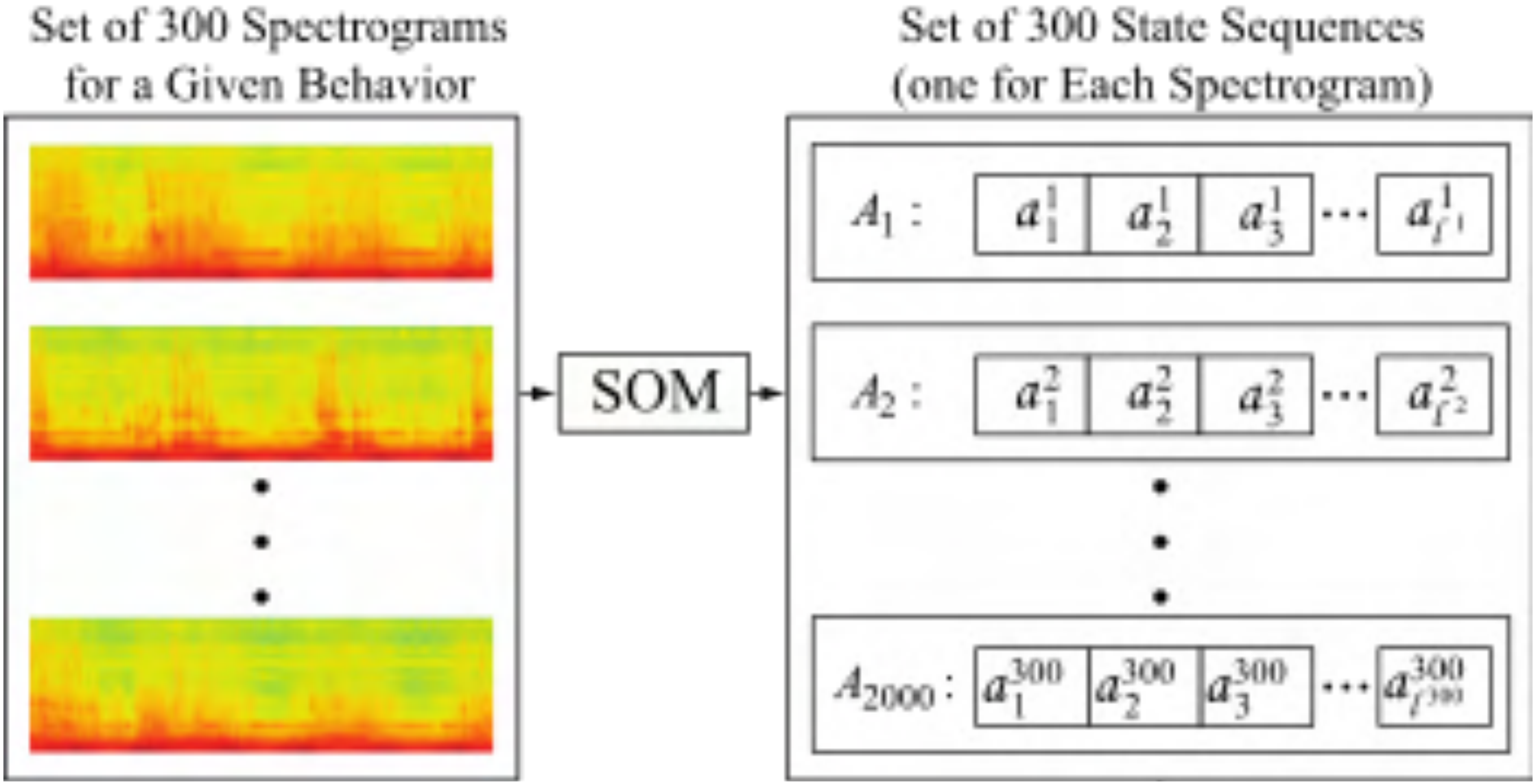
Self Organizing Map:



State Sequence:

A_i: (3,5) -> (2,5) -> (4,3) -> ….

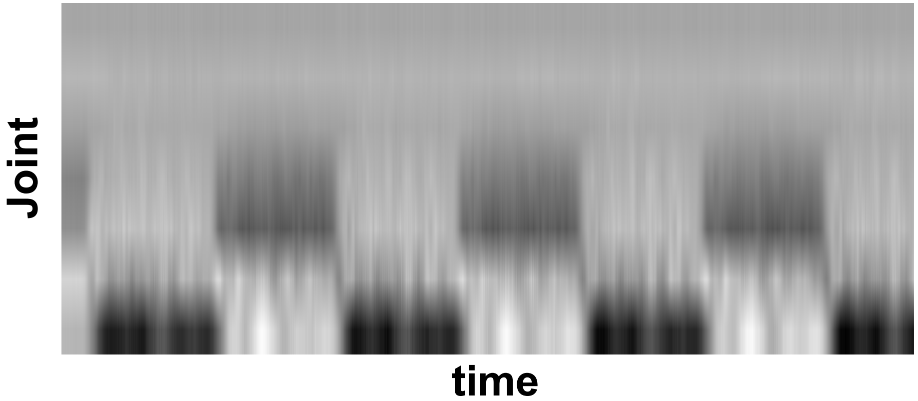
Acoustic Feature Extraction



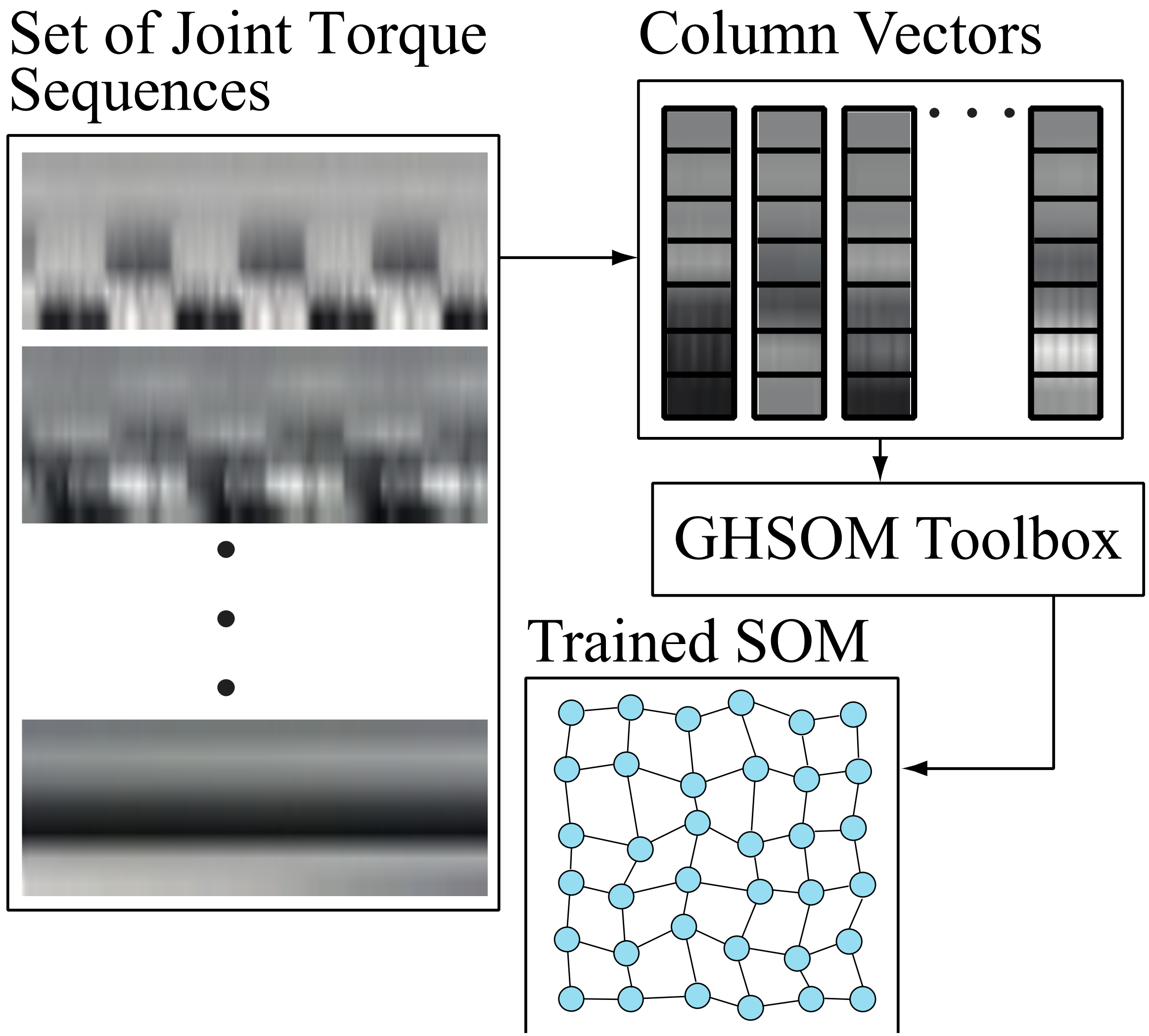
Proprioceptive Feature Extraction

Behavior Execution:
(in and out)

Joint Torque Sequence:

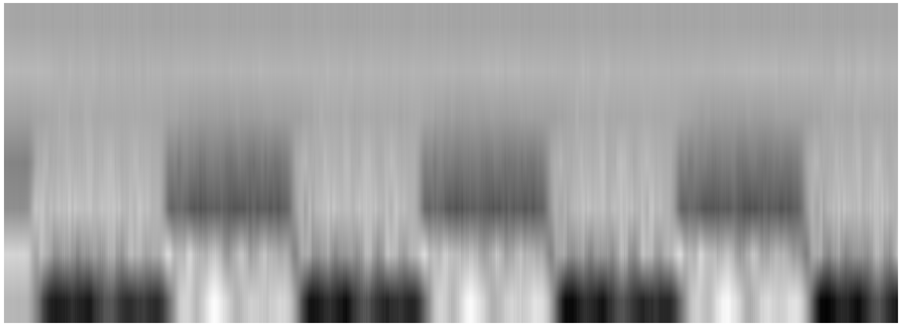


Unsupervised Feature Extraction

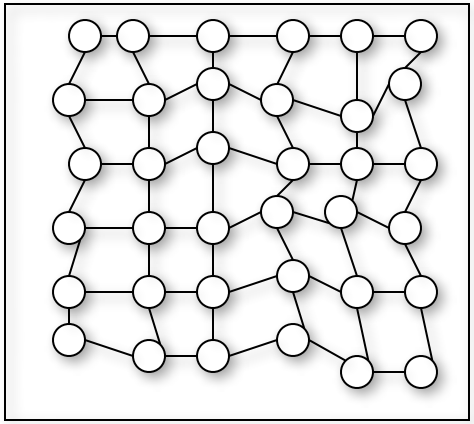


Convert the Joint Torque Sequence to a State Sequence Using a SOM

Joint Torque Sequence:



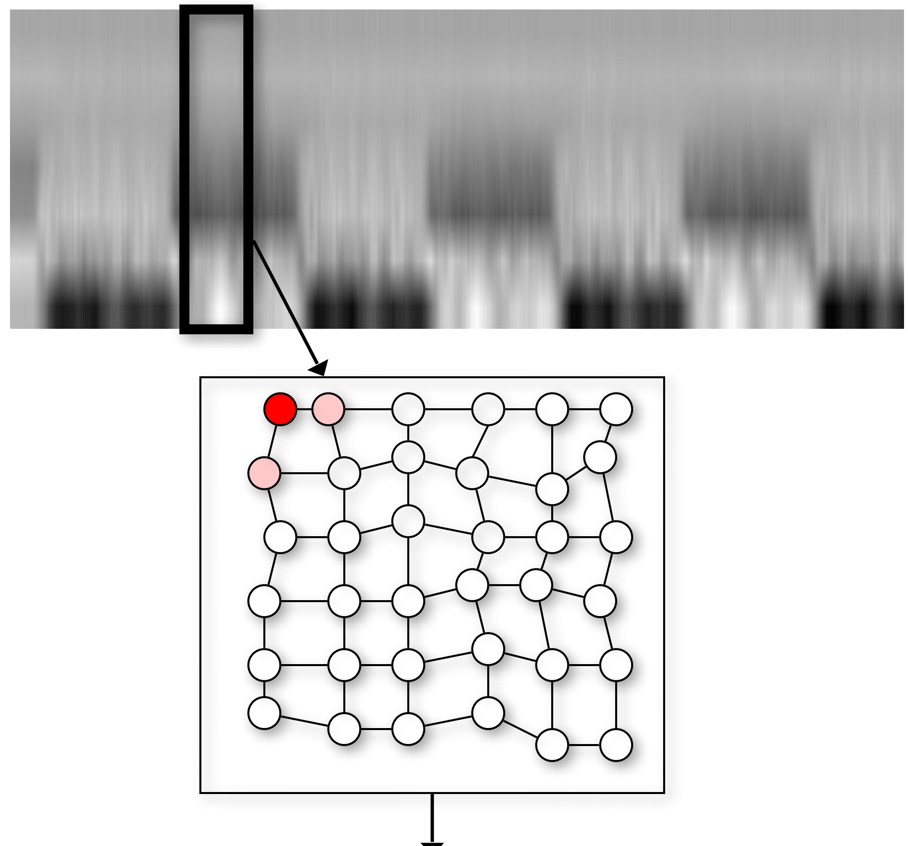
Self Organizing Map:



Convert the Joint Torque Sequence to a State Sequence Using a SOM

Joint Torque Sequence:

Self Organizing Map:



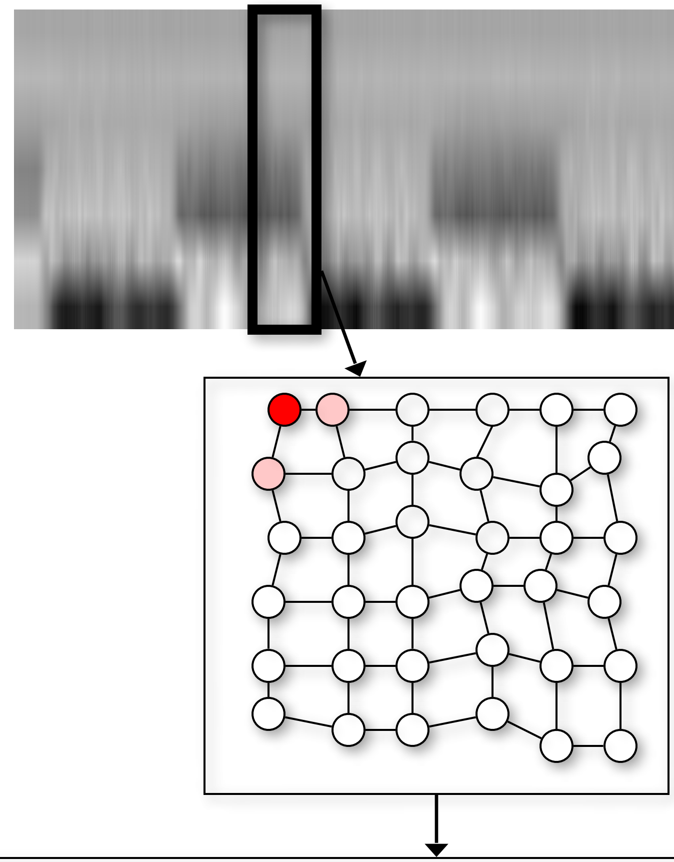
State Sequence:

P_i: (1,6) ->

Convert the Joint Torque Sequence to a State Sequence Using a SOM

Joint Torque Sequence:

Self Organizing Map:

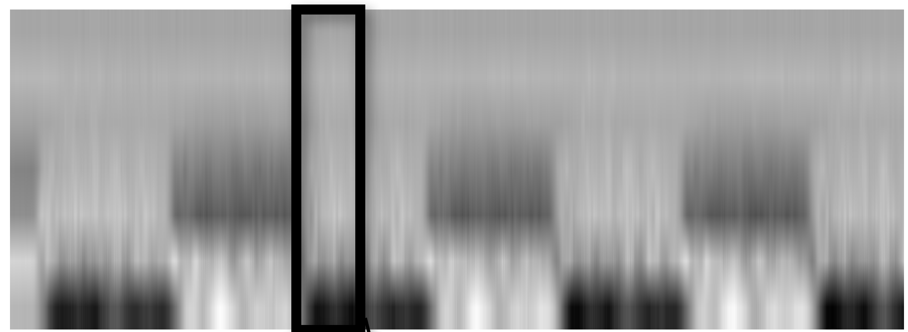


State Sequence:

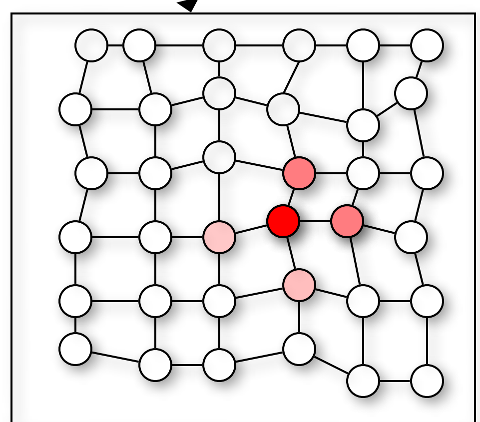
***P_i*: (1,6) -> (1,6) ->**

Convert the Joint Torque Sequence to a State Sequence Using a SOM

Joint Torque Sequence:



Self Organizing Map:



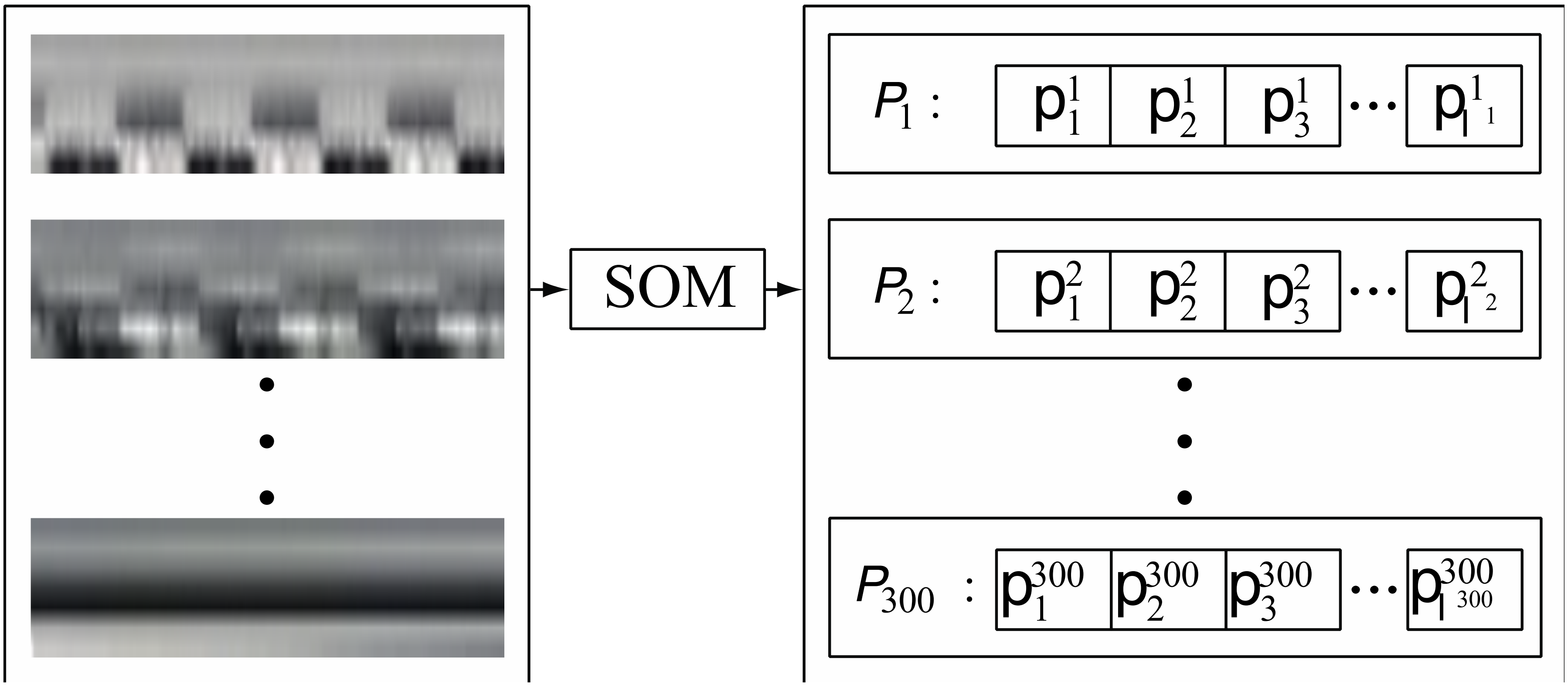
State Sequence:

***P_i*:** **(1,6) -> (1,6) -> (4,3) -> ….**

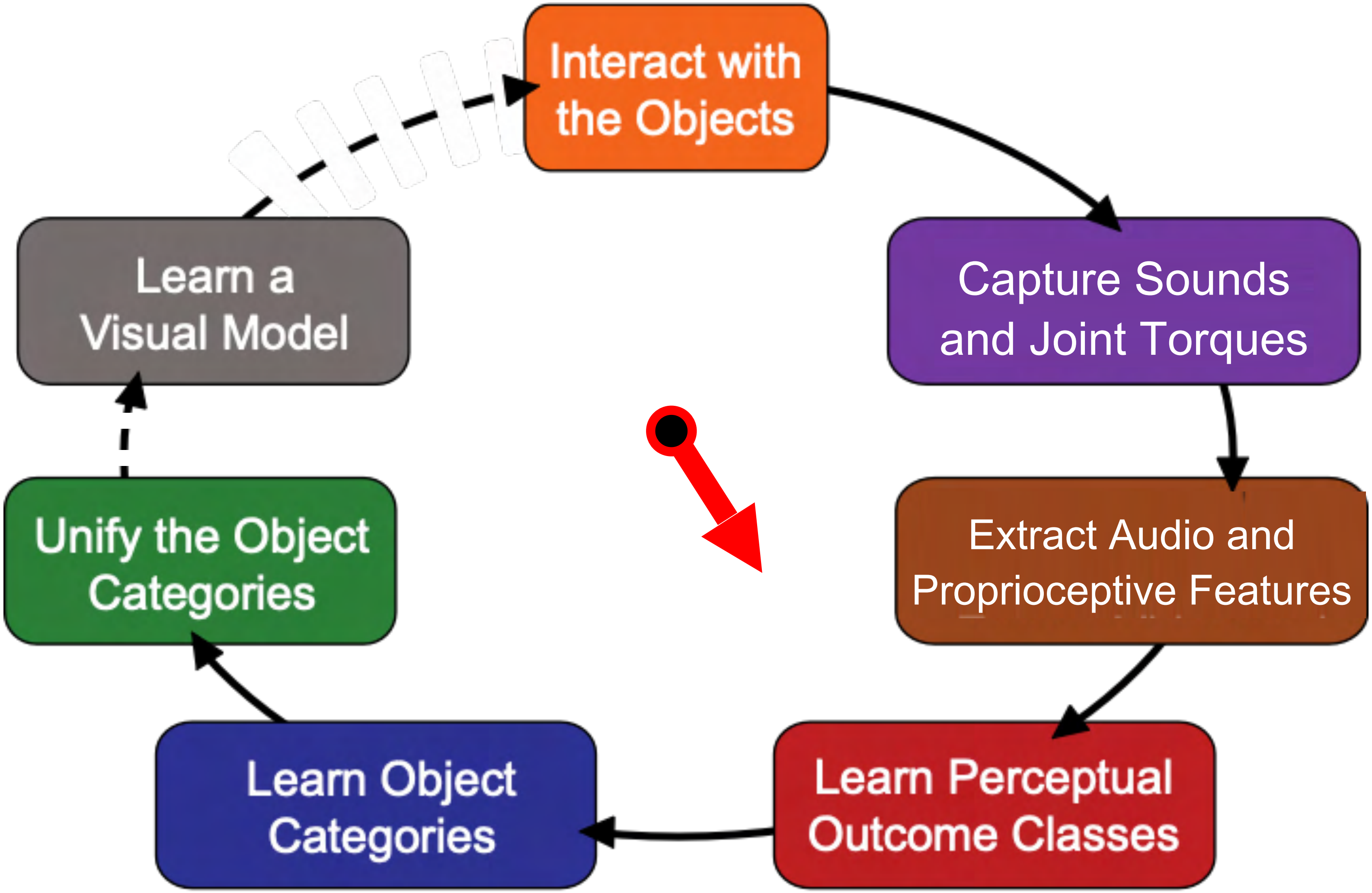
Proprioception Feature Extraction

Set of 300 Joint Torque Sequences for a Given Behavior

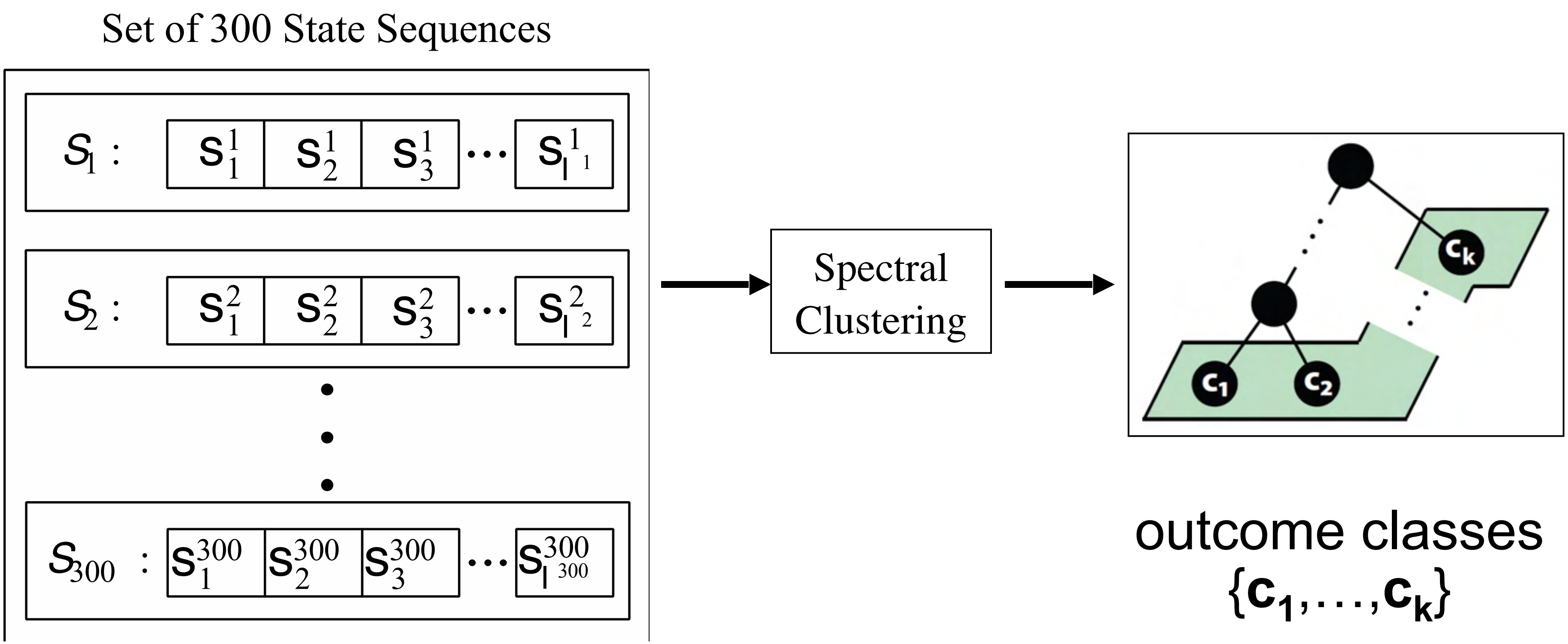
Set of 300 State Sequences (one for Each Joint Torque Sequence)



Learning Framework



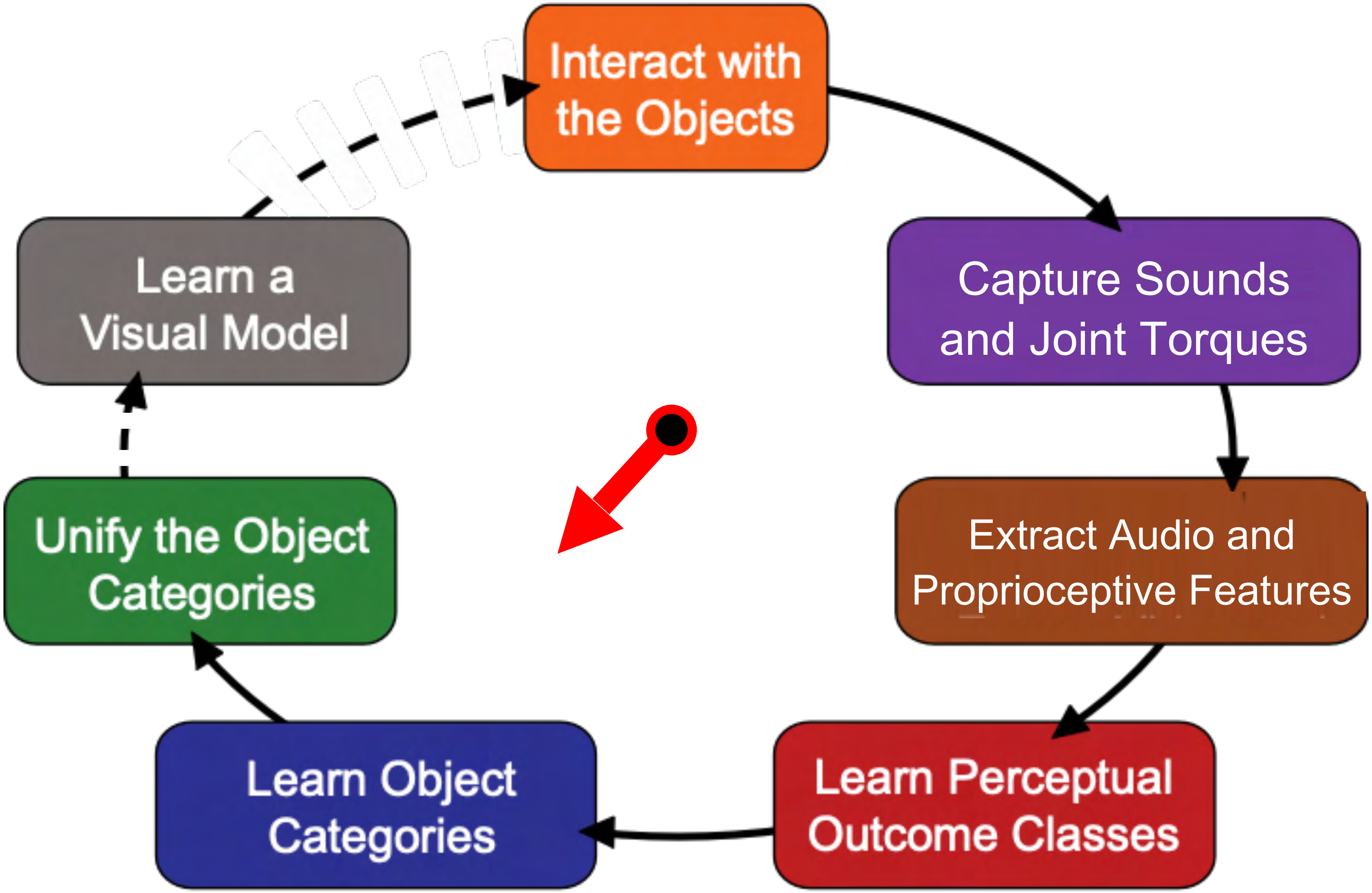
Learning Outcome Classes



- Spectral Clustering requires a similarity matrix as input.
- Similarity function: the Needleman-Wunsch algorithm.

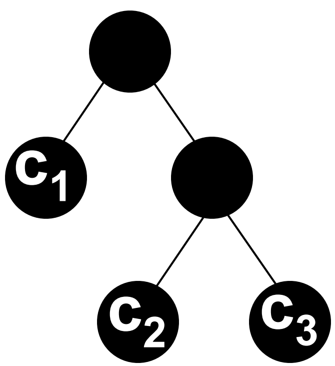
(Needleman and Wunsch, 1970)

Learning Framework



Example Outcome Classes for the In and Out Behavior

Outcome Hierarchy



Learned Outcome Classes

- c_1 Sounds of the water filling up a tall glass
- c_2 Sounds of the water filling up a short cup
- c_3 Sounds of the water splashing against a non-container

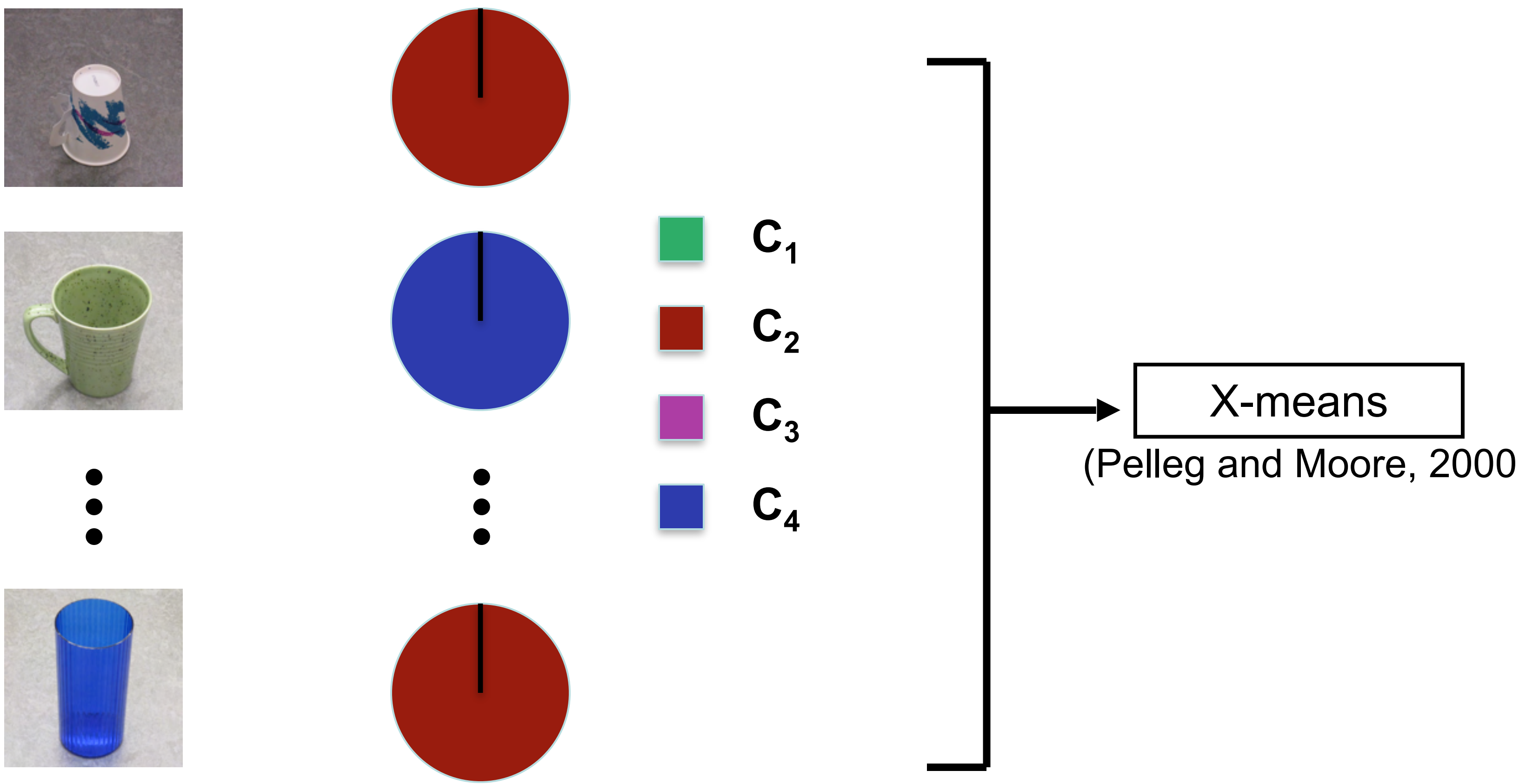
Example Object Representation
(After 10 Repetitions of the In and Out Behavior)

Object	In and Out	Number of Outcome Class Occurrences	Object Representation
		c_1 : 0 c_2 : 0 c_3 : 10	[.0, .0, 1.0]
		c_1 : 10 c_2 : 0 c_3 : 0	[1.0, .0, .0]

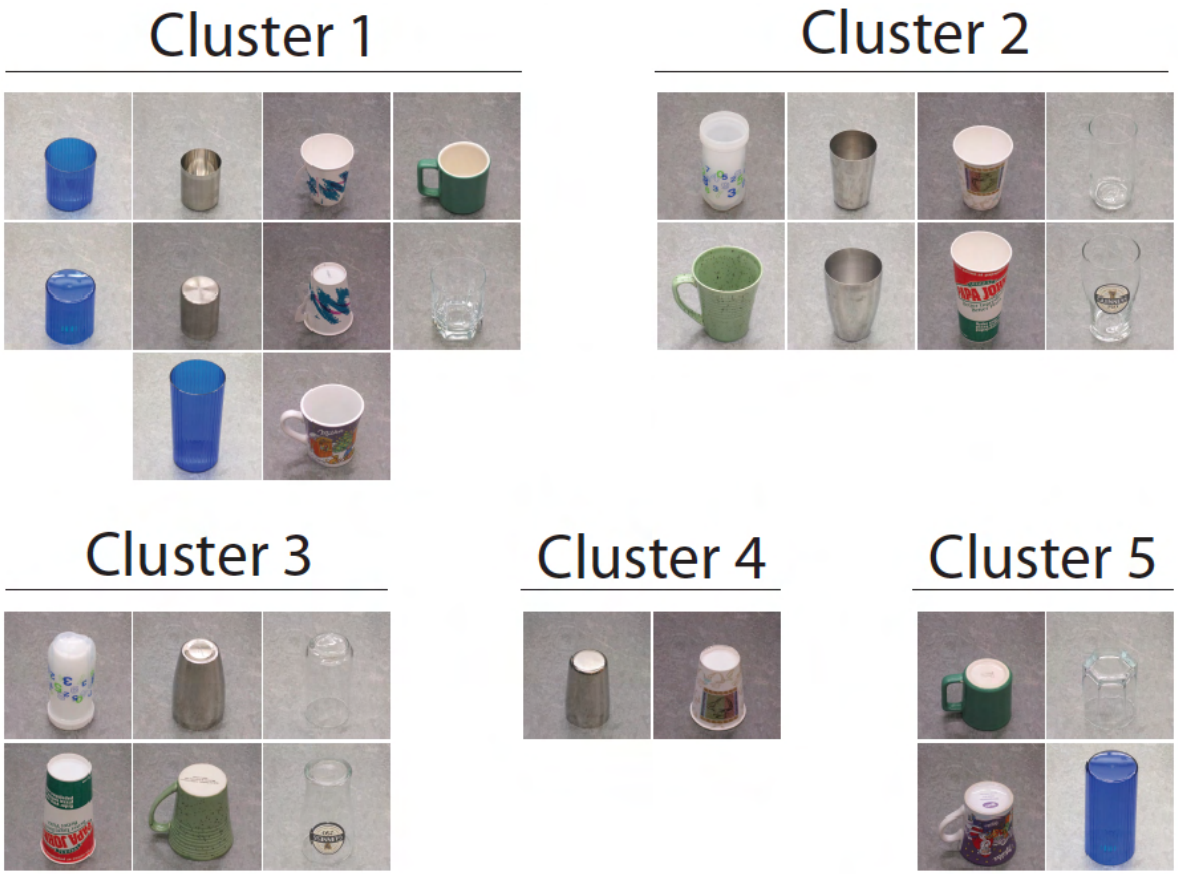
Real Object Representation

(After 300 Repetitions of the In and Out Behavior)

Object | **Number of Outcome Class Occurrences** | **Object Categorization**



Categorization Results



Audio/In and Out

Categorization Results

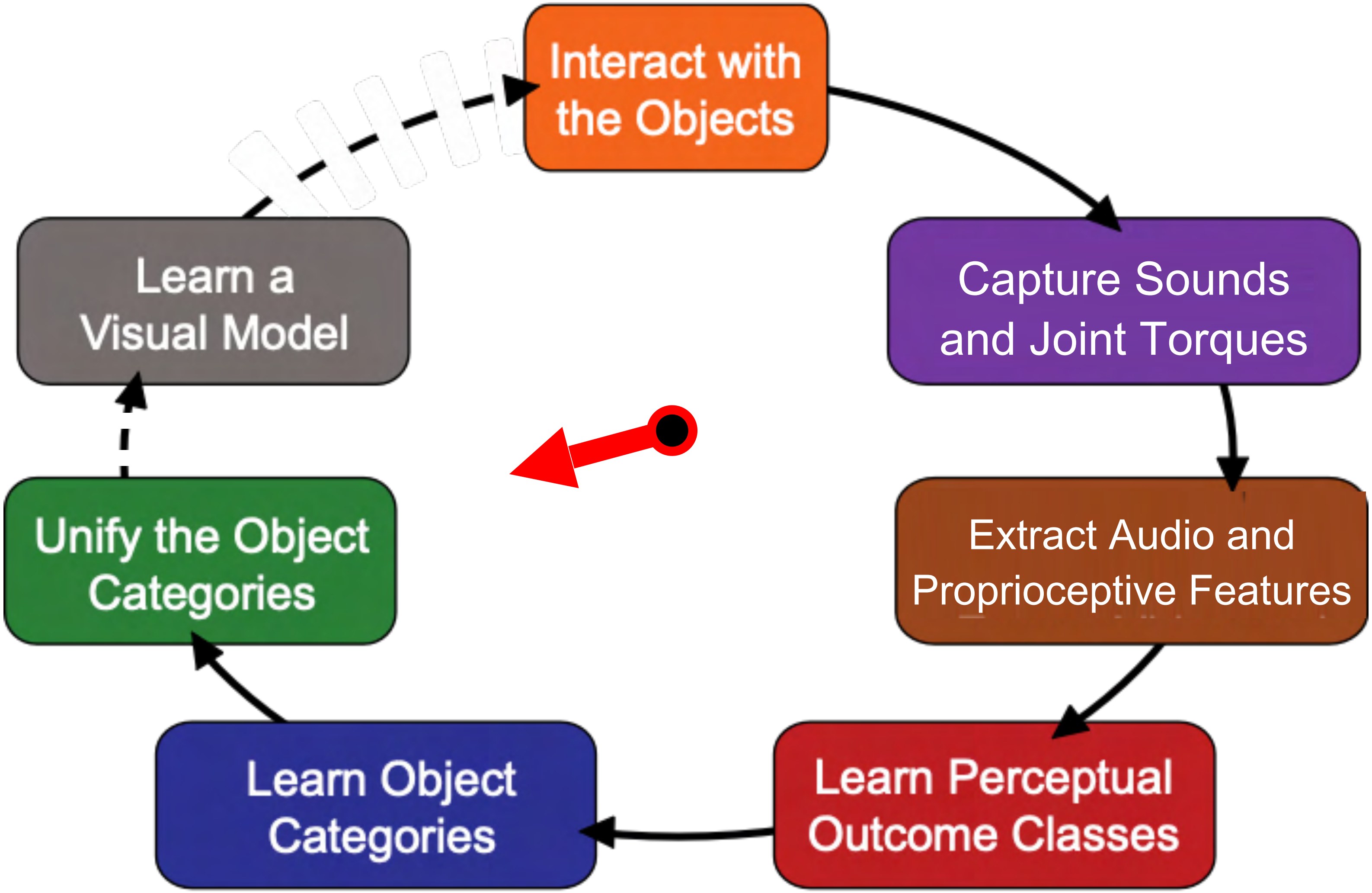
Cluster 1

Cluster 2

Cluster 3

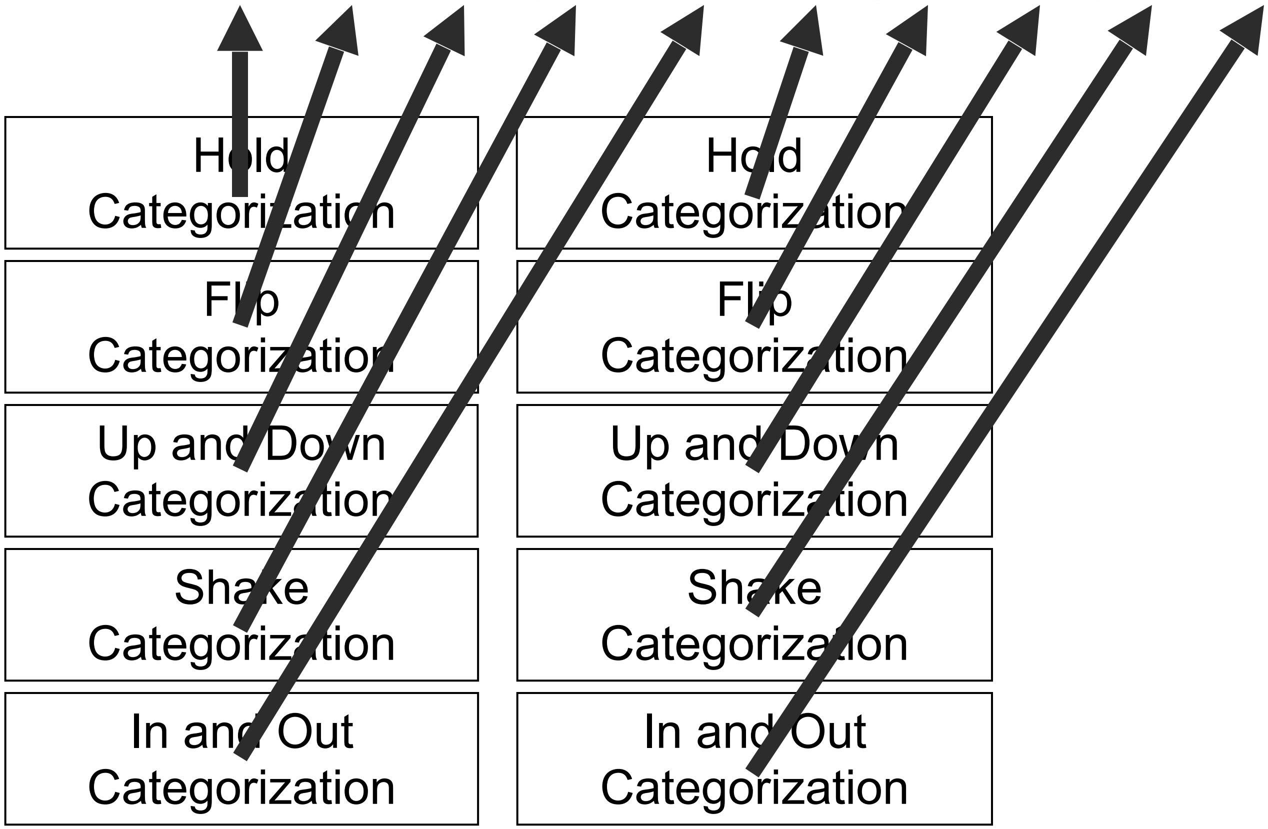
Proprioception/In and Out

Learning Framework



Unification Algorithm

Clusters $\Lambda = \lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)}, \lambda^{(4)}, \lambda^{(5)}, \lambda^{(6)}, \lambda^{(7)}, \lambda^{(8)}, \lambda^{(9)}, \lambda^{(10)}$



Acoustic Clusters **Proprioceptive Clusters**

Unification Algorithm

Clusters

$$\Lambda = \lambda^{(1)}, \ldots, \lambda^{(10)}$$

Hard Consensus Clustering

(Strehl and Ghosh, 2002)

Unified Categorization

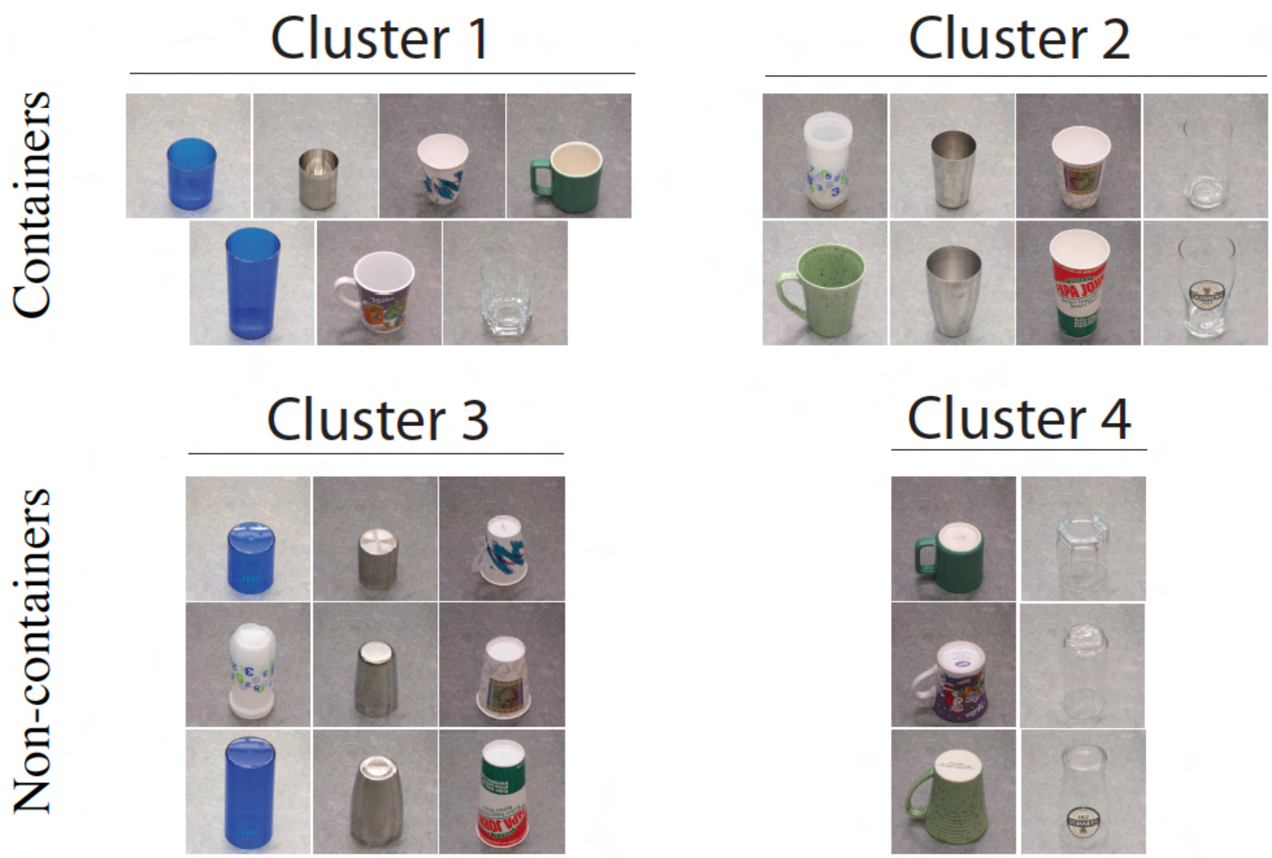
Best found clustering of the objects

- Hard Consensus Clustering searches for a good clustering.

- The output clustering optimizes the normalized mutual information objective function.

Unified Categorization

(derived from both sound and joint torque observations)



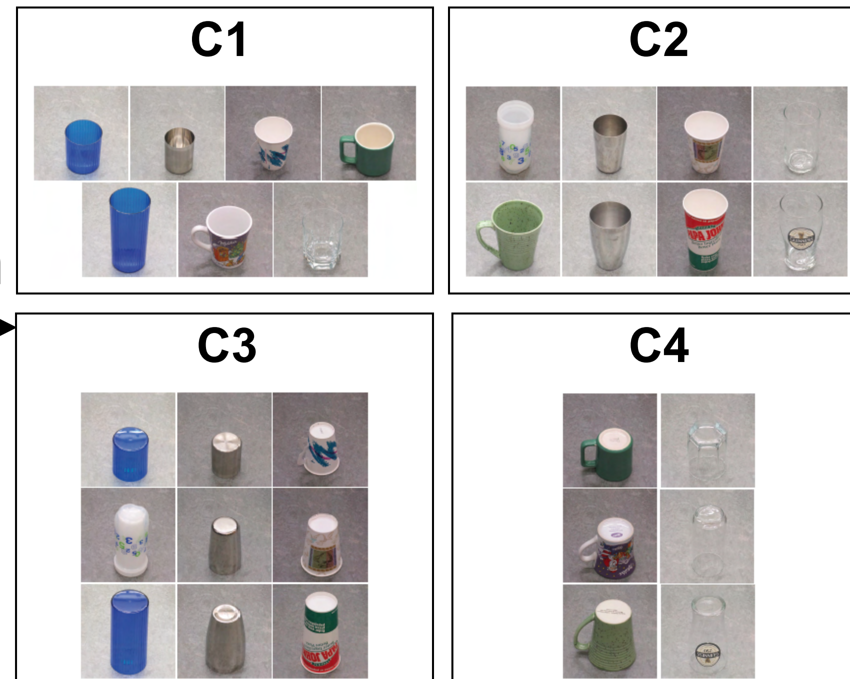
How to Evaluate the Quality of the Categorization?

- Information gain: the change in entropy from a prior state
- entropy of the prior state minus the sum of entropy computed for each cluster

UNCATEGORIZED

information gain

UNIFIED CLUSTERING



Categorization Quality

UNIFIED CLUSTERING

Cluster	Conditional Probabilities: container	Conditional Probabilities: non-container	Entropy	Information Gain
C1 	7/7	0/7		
C2 	8/8	0/8		
C3 	0/9	9/9		
C4 	0/6	6/6		

Categorization Quality

UNIFIED CLUSTERING

Cluster	Conditional Probabilities		Entropy	Information Gain
	container	non-container		
C1 	7/7	0/7	→	$-(7/7)\log_2(7/7) - (0/7)\log_2(0/7)$
C2 	8/8	0/8	→	$-(8/8)\log_2(8/8) - (0/8)\log_2(0/8)$
C3 	0/9	9/9	→	$-(0/9)\log_2(0/9) - (9/9)\log_2(9/9)$
C4 	0/6	6/6	→	$-(0/6)\log_2(0/6) - (6/6)\log_2(6/6)$

Categorization Quality

UNIFIED CLUSTERING

Cluster	Conditional Probabilities: container	Conditional Probabilities: non-container	Entropy	Information Gain
C1 	7/7	0/7	→ 0.0	
C2 	8/8	0/8	→ 0.0	
C3 	0/9	9/9	→ 0.0	
C4 	0/6	6/6	→ 0.0	

Categorization Quality

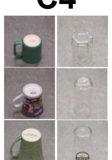
UNIFIED CLUSTERING

Cluster	Conditional Probabilities: container	Conditional Probabilities: non-container		Entropy	Information Gain
C1 	7/7	0/7	→	0.0	– (7/30 * 0.0)
C2 	8/8	0/8	→	0.0	– (8/30 * 0.0)
C3 	0/9	9/9	→	0.0	– (9/30 * 0.0)
C4 	0/6	6/6	→	0.0	– (6/30 * 0.0)

1 – (7/30 * 0.0) – (8/30 * 0.0) – (9/30 * 0.0) – (6/30 * 0.0)

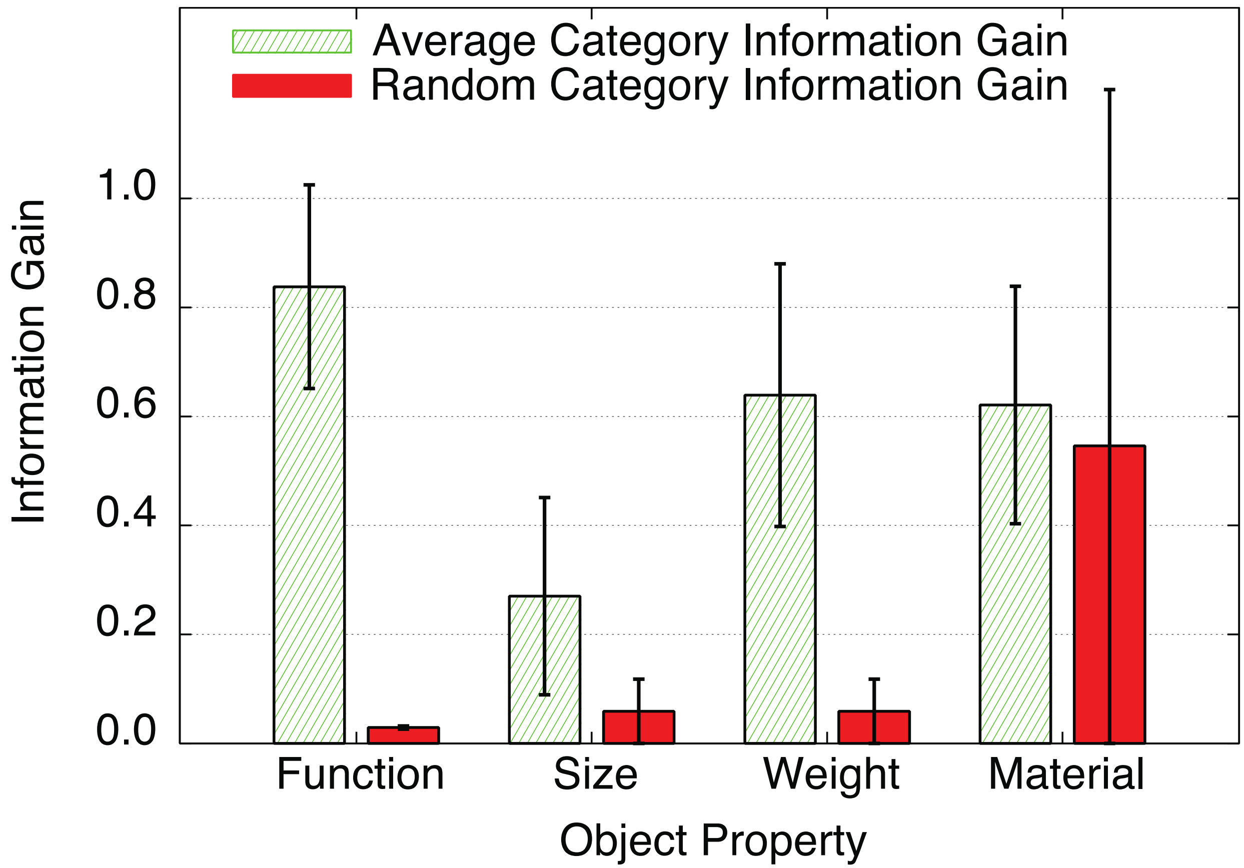
Categorization Quality

UNIFIED CLUSTERING

Cluster	Conditional Probabilities: container	Conditional Probabilities: non-container		Entropy	Information Gain
C1 	7/7	0/7	→	0.0	– (7/30 * 0.0)
C2 	8/8	0/8	→	0.0	– (8/30 * 0.0)
C3 	0/9	9/9	→	0.0	– (9/30 * 0.0)
C4 	0/6	6/6	→	0.0	– (6/30 * 0.0)

1.0

Information Gained with Respect to Human Labels



Conclusion

- Object category learning is possible while interacting with objects in a sink.
- The sounds and the sensations of water flowing into a cup is one embodiment of that object.
- Sound and proprioception might be able to bootstrap water manipulation research.

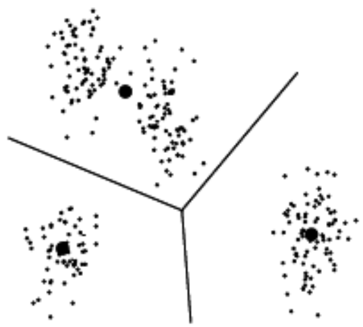
Acknowledgements

Questions?

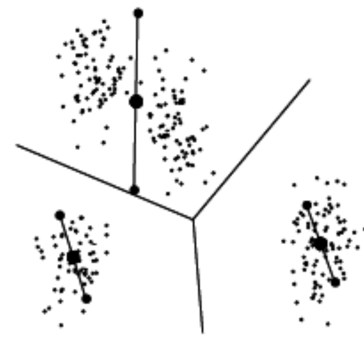
Material Properties of the Objects

X-means (Pelleg and Moore, 2000)

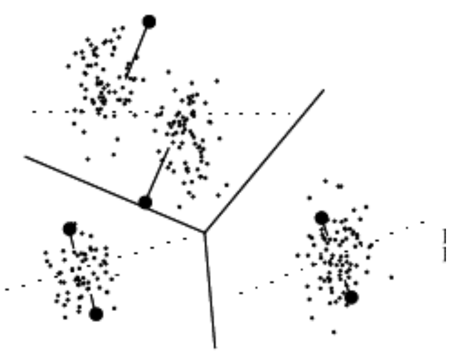
K-means Result



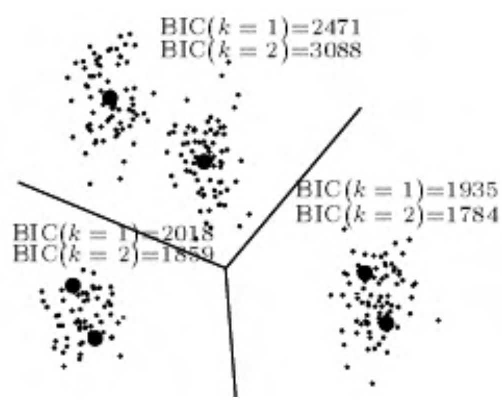
Split the Centroids



Run K-means locally



Evaluate the New Centroids



New Structure

Needleman-Wunsch (1970)

Initialize the Matrix

		C	O	E	L	A	C	A	N	T	H
	0	← -1	← -2	← -3	← -4	← -5	← -6	← -7	← -8	← -9	← -10
P	↑ -1										
E	↑ -2										
L	↑ -3										
I	↑ -4										
C	↑ -5										
A	↑ -6										
N	↑ -7										

Fill the Matrix

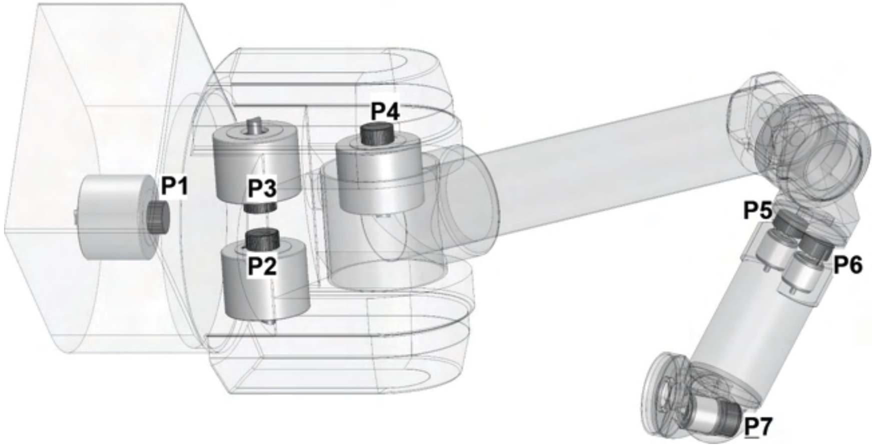
		C	O	E	L	A	C	A	N	T	H
	0	← -1	← -2	← -3	← -4	← -5	← -6	← -7	← -8	← -9	← -10
P	↑ -1	↖ **-1**	↖ -2	↖ -3	↖ -4	↖ -5	↖ -6	↖ -7	↖ -8	↖ -9	↖ -10
E	↑ -2	↖ -2	↖ -2	↖ -1	← -0	← -3	← -4	← -5	← -6	← -7	← -8
L	↑ -3	↖ -3	↖ -3	← -2	↖ -2	← -1	← -2	← -3	← -4	← -5	← -6
I	↑ -4	↖ -4	↑ -4	↑ -3	↑ -1	↖ -1	↖ -2	↖ -1	↖ -4	↖ -5	↖ -6
C	↑ -5	↖ -3	← -4	↑ -4	↑ -2	↖ -2	↖ -0	← -1	← -2	← -3	← -4
A	↑ -6	↑ -4	↖ -4	↖ -5	↑ -3	↖ -1	↑ -1	↖ -1	← -0	← -1	← -2
N	↑ -7	↖ -5	↖ -5	↖ -5	↑ -4	↑ -2	↖ -2	← -0	↖ -2	← -1	← -0

What Sensory Modalities to Use?

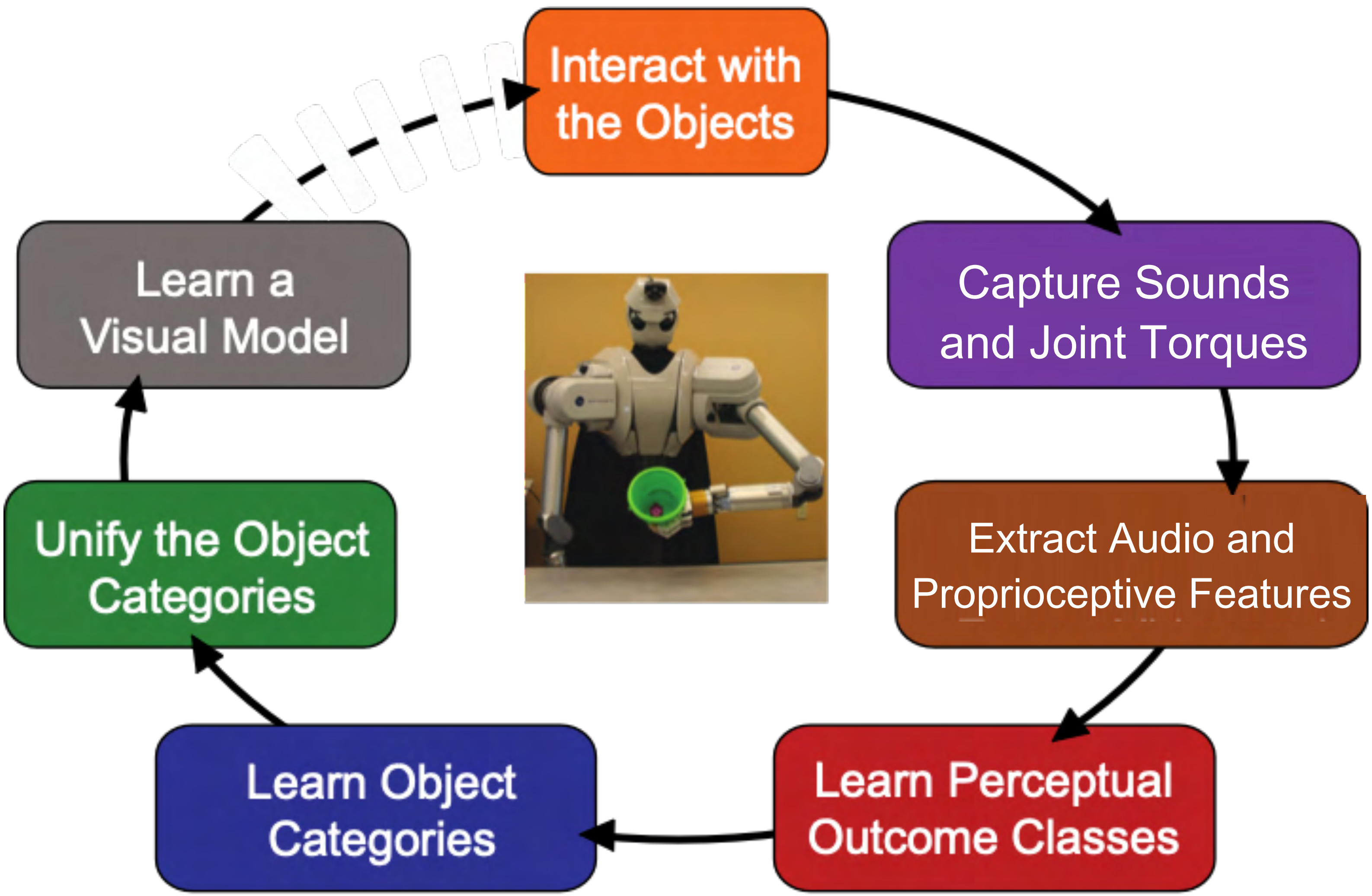
Vision

Audio

Proprioception



Learning Framework



Learning Framework