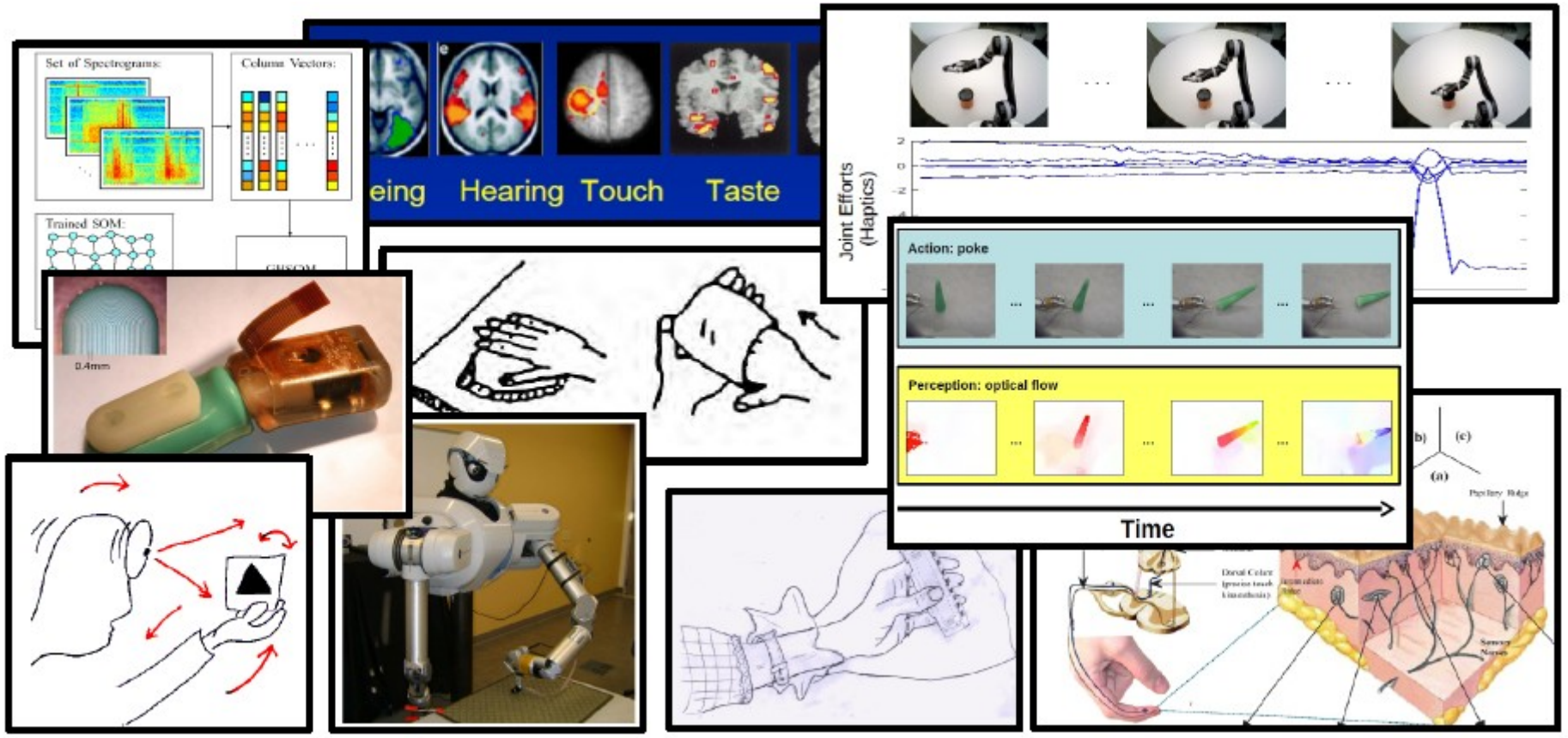




COMP 141: Probabilistic Robotics for Human-Robot Interaction

Instructor: Jivko Sinapov
www.cs.tufts.edu/~jsinapov

This week: Perception beyond Vision

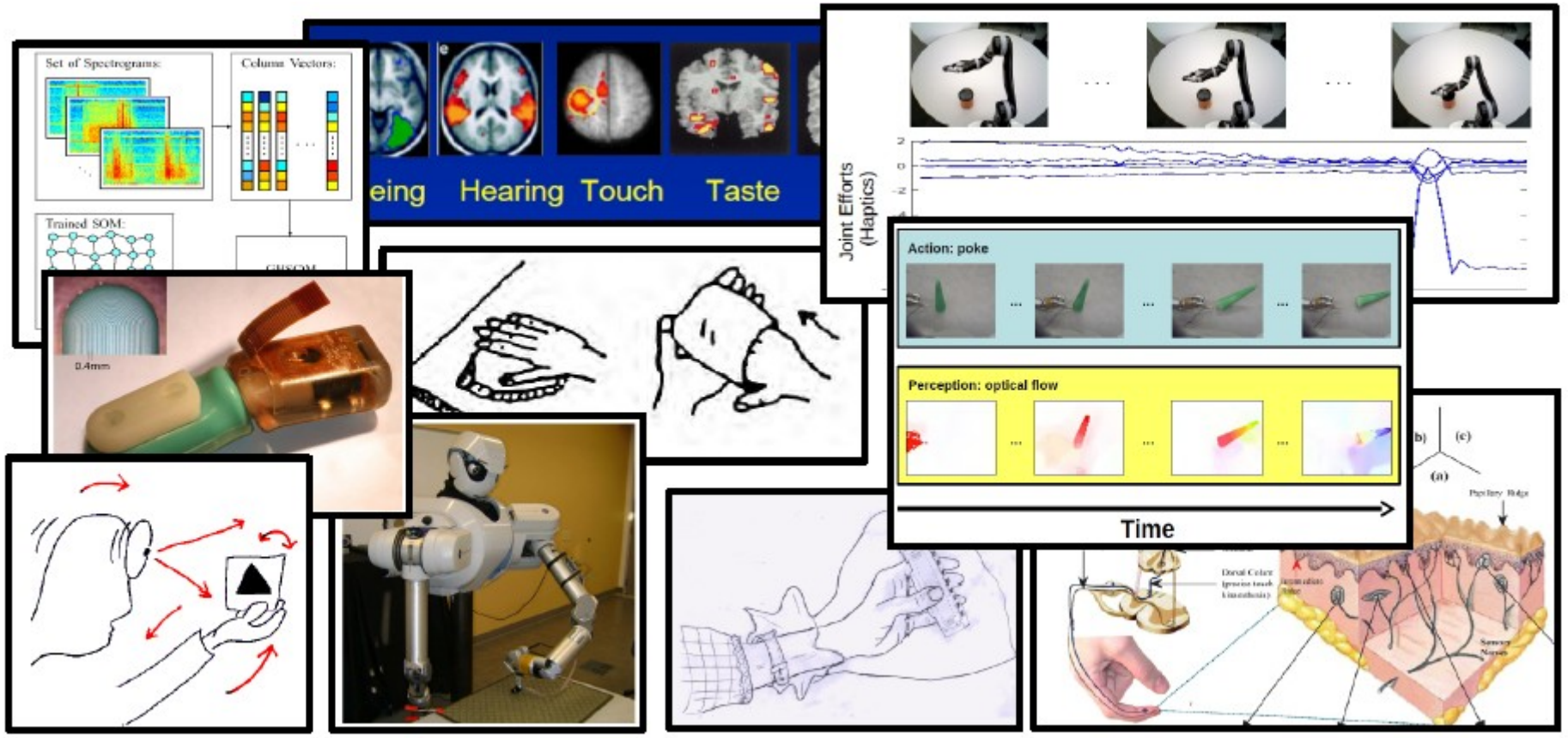


Announcements

Research Article Reading Assignment #3

- Due next Tuesday

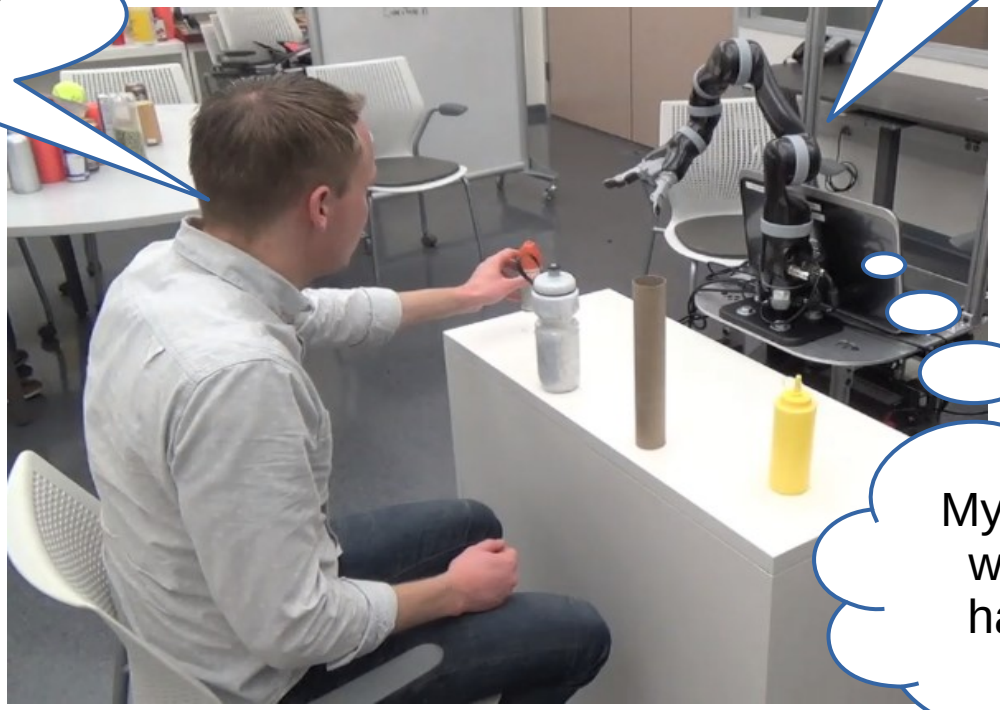
Today: Perception beyond Vision



Language Acquisition

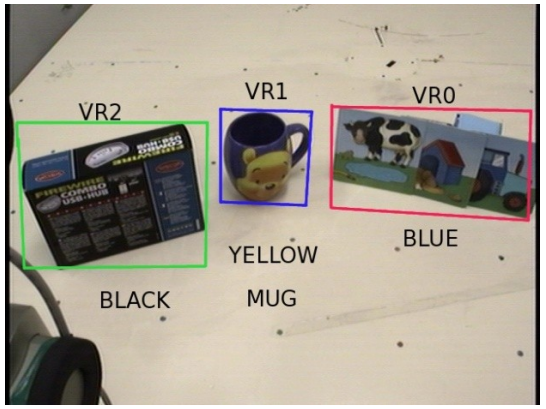
How would you describe this object?

It is a small orange spray can

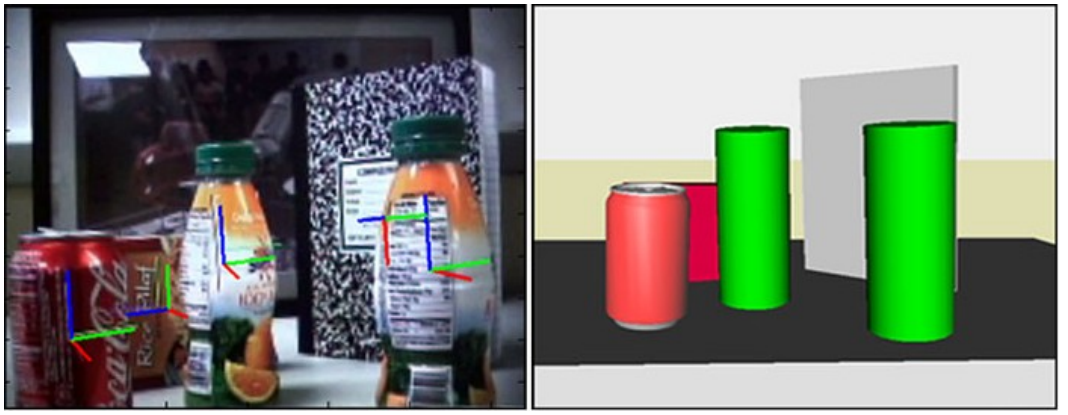


My model of the word ‘orange’ has improved!

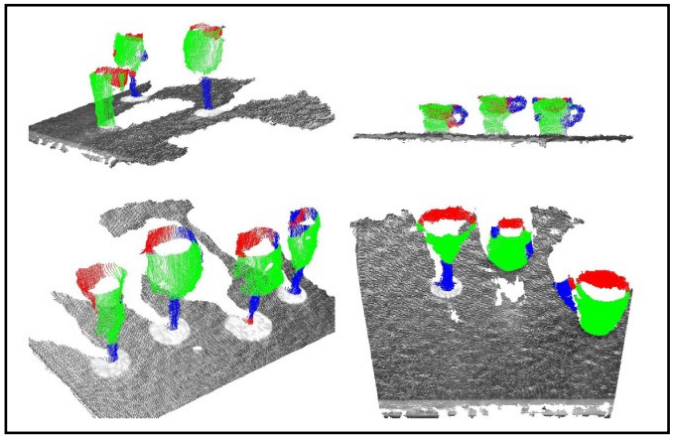
Current Solution: connect the symbol with visual input



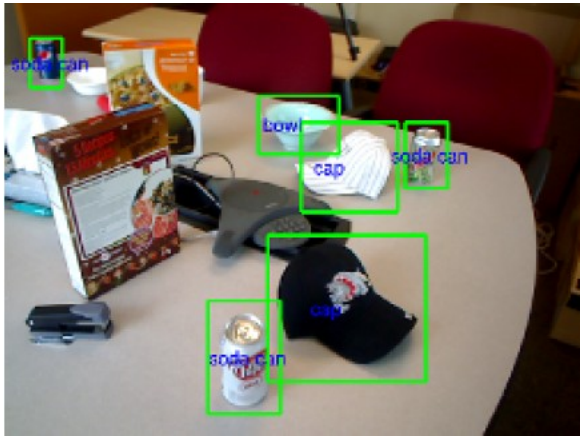
Sridharan *et al*. 2008



Collet *et al*. 2009

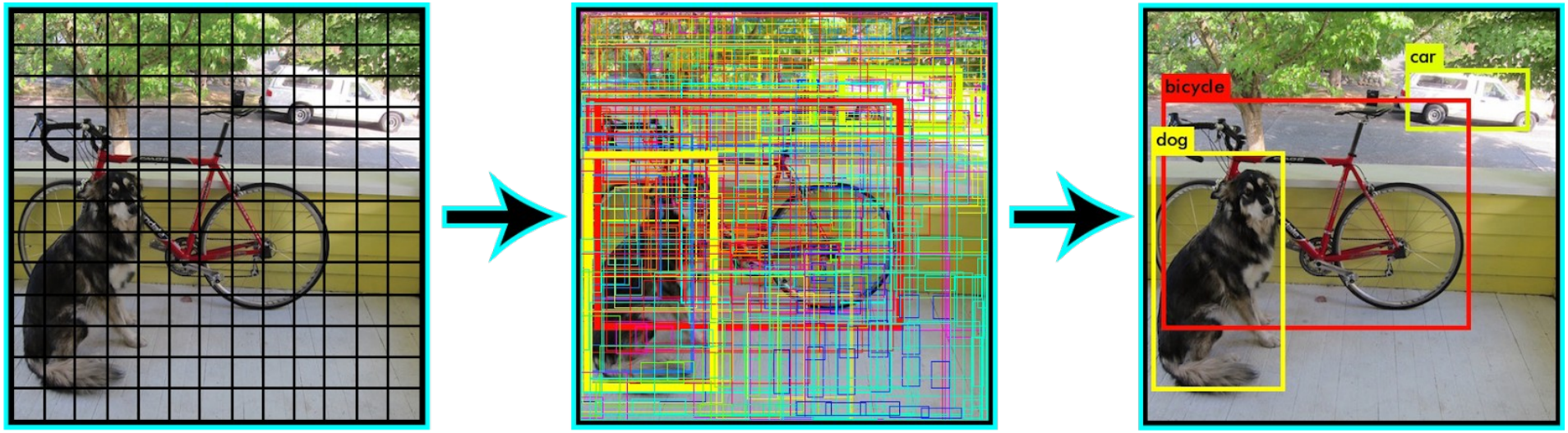


Rusu *et al*. 2009



Lai *et al*. 2011

Current Solution: connect the symbol with visual input



Redmon *et al*. 2016

“Robot, I am thirsty, fetch me the ***yellow juice carton***”

“Robot, I am thirsty, fetch me the ***full juice carton***”

Solution: Lift the Object

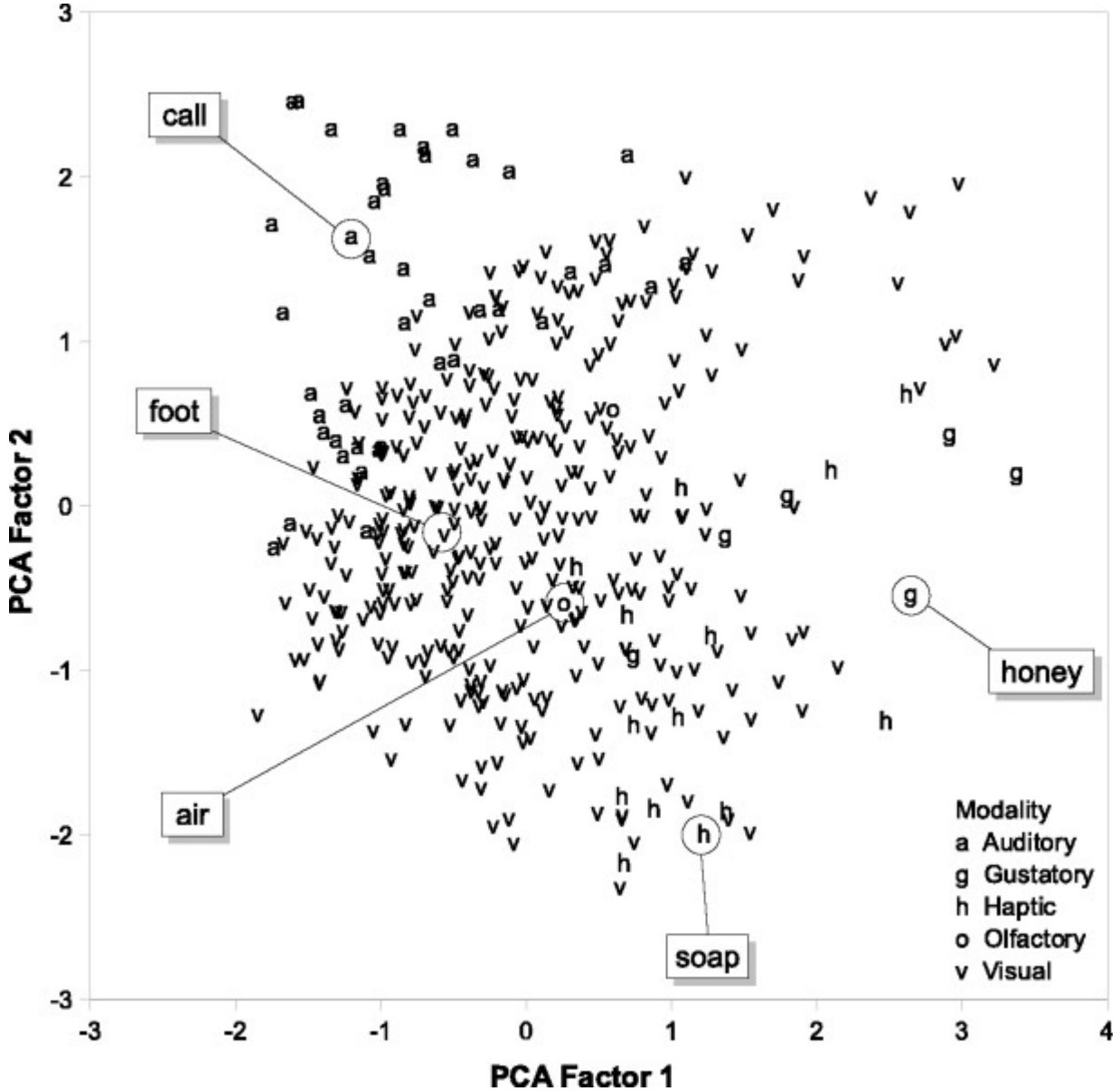
“Fetch me the ***pill bottle***”

Solution: Shake the Object

Solution: Shake the Object

Exploratory behaviors give us information about objects that vision cannot!

Modality Exclusivity Norms for common English nouns and adjectives



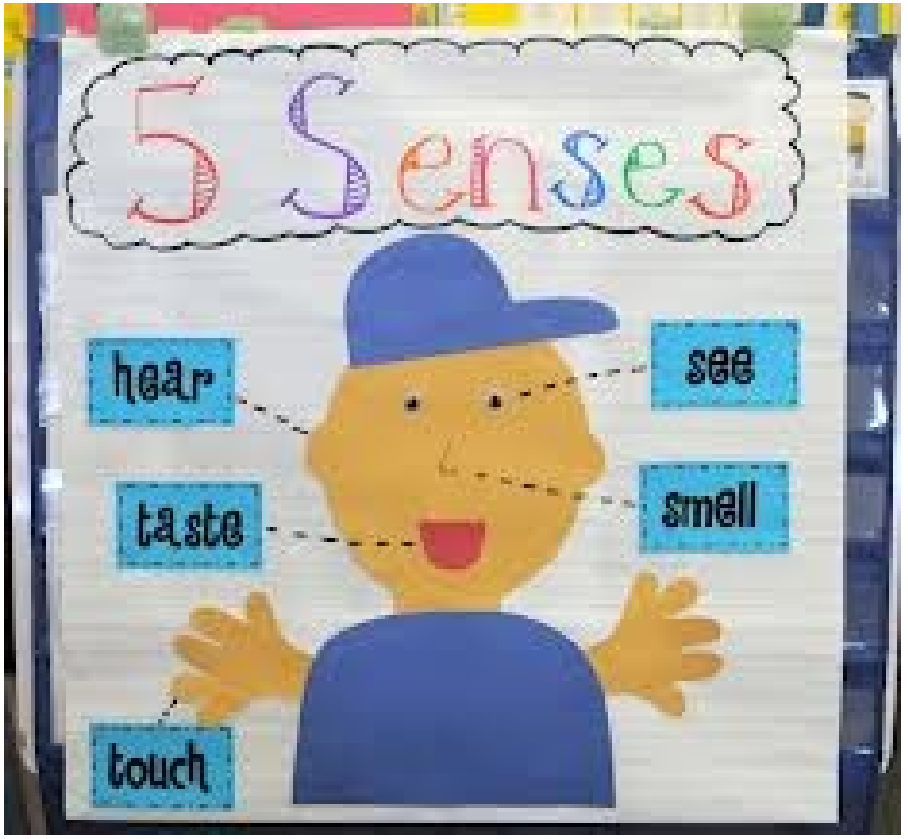
KNOWLEDGE ABOUT OBJECT	EXPLORATORY PROCEDURE
Substance-related properties	
Texture	Lateral motion
Hardness	Pressure
Temperature	Static contact
Weight	Unsupported holding
Structure-related properties	
Weight	Unsupported holding
Volume	Enclosure, contour following
Global shape	Enclosure
Exact shape	Contour following
Functional properties	
Part motion	Part motion test
Specific motion	Function test

[Power, 2000]

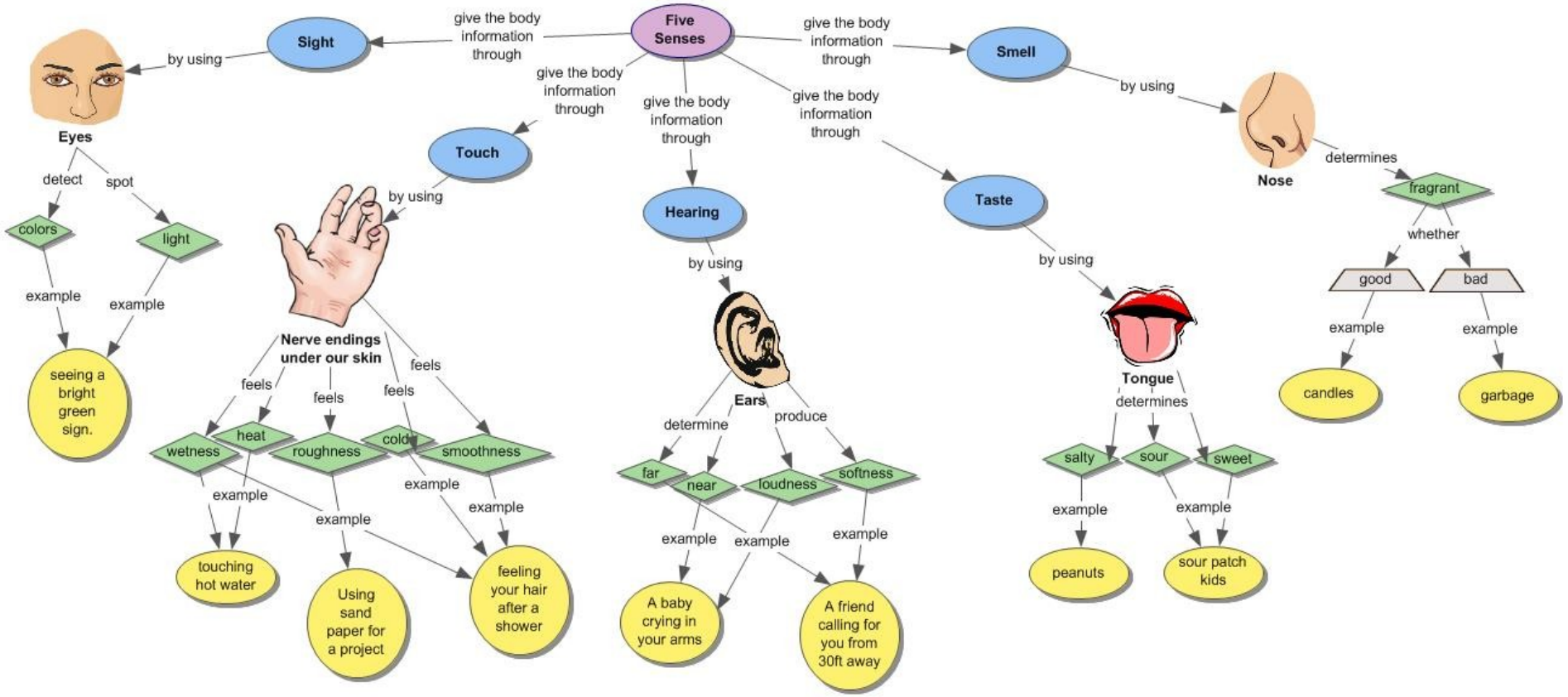
[Lederman and Klatzky, 1987]

Object Exploration in Infancy

The “5” Senses



The “5” Senses



[http://edublog.cmich.edu/meado1bl/files/2013/03/Five-Senses2.jpg]

MAKING SENSE OF THE SENSES

There are many opinions about how many senses we have

SENSORY MODALITY	Conservative	Accepted	Radical
Vision	■		
Light		■	■
Colour		■	
Red			■
Green			■
Blue			■
Hearing	■	■	■
Smell	■	■	
2000 or more receptor types			■
Taste	■		
Sweet		■	■
Salt		■	■
Sour		■	■
Bitter		■	■
Umami			■
Touch	■	■	
Light touch			■
Pressure			■
Pain	■	■	
Cutaneous			■
Somatic			■
Visceral			■
Mechanoreception	■		
Balance	■	■	
Rotational acceleration			■
Linear acceleration			■
Proprioception – joint position		■	■
Kinaesthesis		■	
Muscle stretch – Golgi tendon organs			■
Muscle stretch – muscle spindles			■
Temperature	■		
Heat		■	■
Cold		■	■
Interoceptors			
Blood pressure	■	■	
Arterial blood pressure			■
Central venous blood pressure			■
Head blood temperature			■
Blood oxygen content		■	■
Cerebrospinal fluid pH		■	■
Plasma osmotic pressure (thirst?)		■	■
Artery-vein blood glucose difference (hunger?)		■	■
Lung inflation		■	■
Bladder stretch			■
Full stomach			■
TOTAL	10	21	33

MAKING SENSE OF THE SENSES

There are many opinions about how many senses we have

SENSORY MODALITY	Conservative	Accepted	Radical
Vision	■		
Light		■	■
Colour		■	
Red			■
Green			■
Blue			■
Hearing	■	■	■
Smell	■	■	
2000 or more receptor types			■
Taste	■		
Sweet		■	■
Salt		■	■
Sour		■	■
Bitter		■	■
Umami			■
Touch	■	■	
Light touch			■
Pressure			■
Pain	■	■	
Cutaneous			■
Somatic			■
Visceral			■
Mechanoreception	■		
Balance	■	■	
Rotational acceleration			■
Linear acceleration			■
Proprioception – joint position		■	■
Kinaesthesis		■	
Muscle stretch – Golgi tendon organs			■
Muscle stretch – muscle spindles			■
Temperature	■		
Heat		■	■
Cold		■	■
Interoceptors			
Blood pressure	■	■	
Arterial blood pressure			■
Central venous blood pressure			■
Head blood temperature			■
Blood oxygen content		■	■
Cerebrospinal fluid pH		■	■
Plasma osmotic pressure (thirst?)		■	■
Artery-vein blood glucose difference (hunger?)		■	■
Lung inflation		■	■
Bladder stretch			■
Full stomach			■
TOTAL	10	21	33

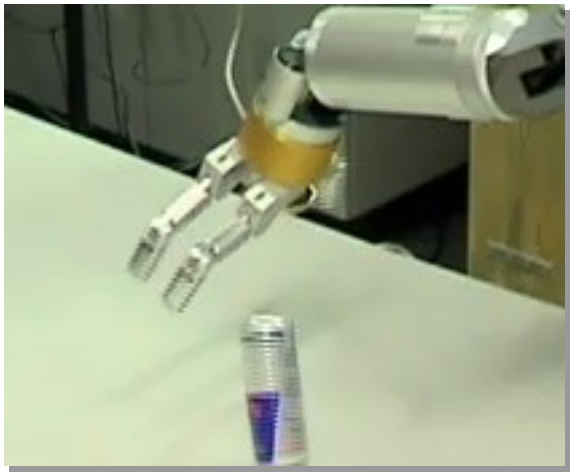
Why sound for robotics?

What just happened?

What just happened?

What actually happened:

The robot dropped a soda-can



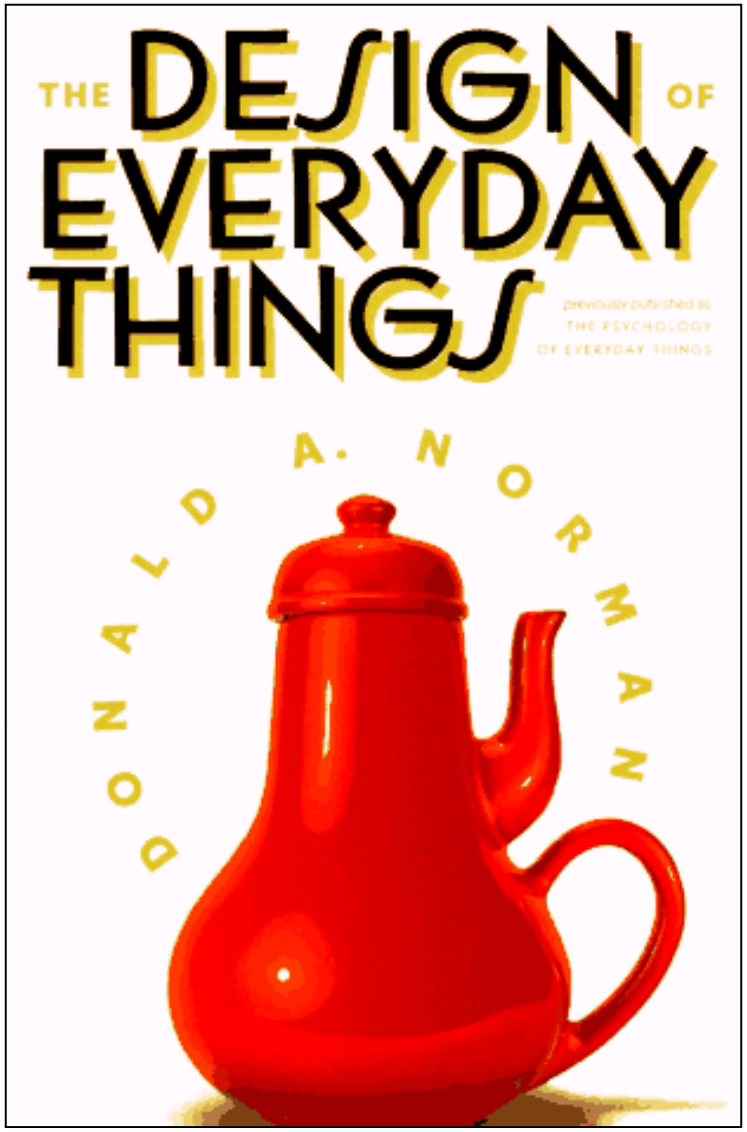
Why Natural Sound is Important

"…natural sound is as essential as visual information because **sound tells us about things that we can't see**, and it does so while our eyes are occupied elsewhere. "

"Sounds are generated when materials interact, and the sounds tell us whether they are **hitting, sliding, breaking, tearing, crumbling, or bouncing**. "

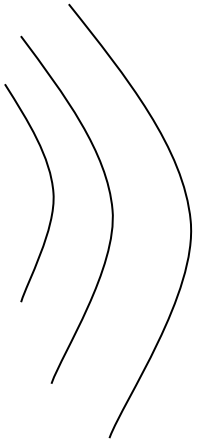
"Moreover, **sounds differ according to the characteristics of the objects**, according to their size, solidity, mass, tension, and material. "

Don Norman, "*The Design of Everyday Things*", p.103



Why Natural Sound is Important

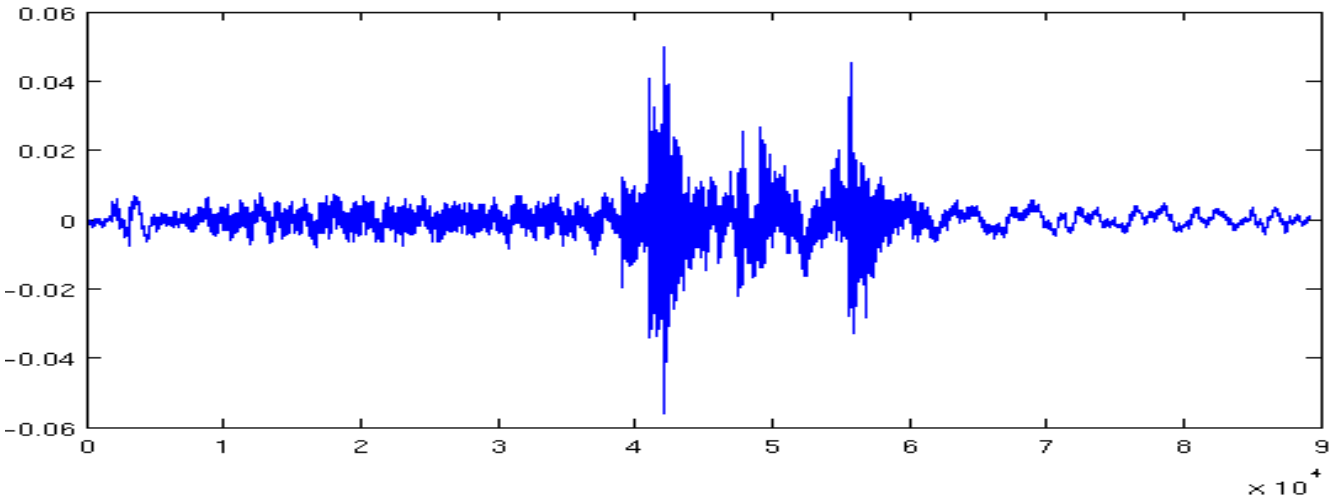
Sound Producing Event



[Gaver, 1993]

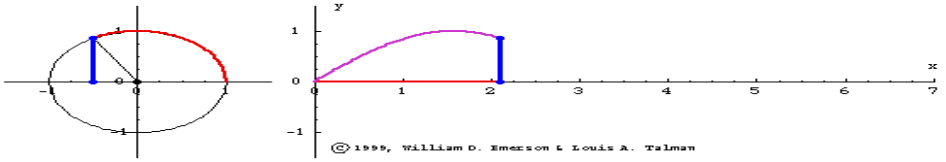
What is a Sound Wave?

What is a Sound Wave?



….from a computer's point of view, raw audio is a sequence of 44.1K floating point numbers arriving each second

Sine Curve



[http://clem.mscd.edu/~talmanl/HTML/SineCurve.html]

Frequency

- Measured in Hertz (Hz)
- Named after Heinrich Hertz
- 1 Hertz = 1 repetition per second
- Typically denoted with the letter f

Period

- How long does one cycle take?
- It is the reciprocal of the frequency
- Measured in seconds
- Typically denoted with the letter T

Frequency vs Period Animation

f = 0.5 Hz
T = 2.0 s

f = 1.0 Hz
T = 1.0 s

f = 2.0 Hz
T = 0.5 s

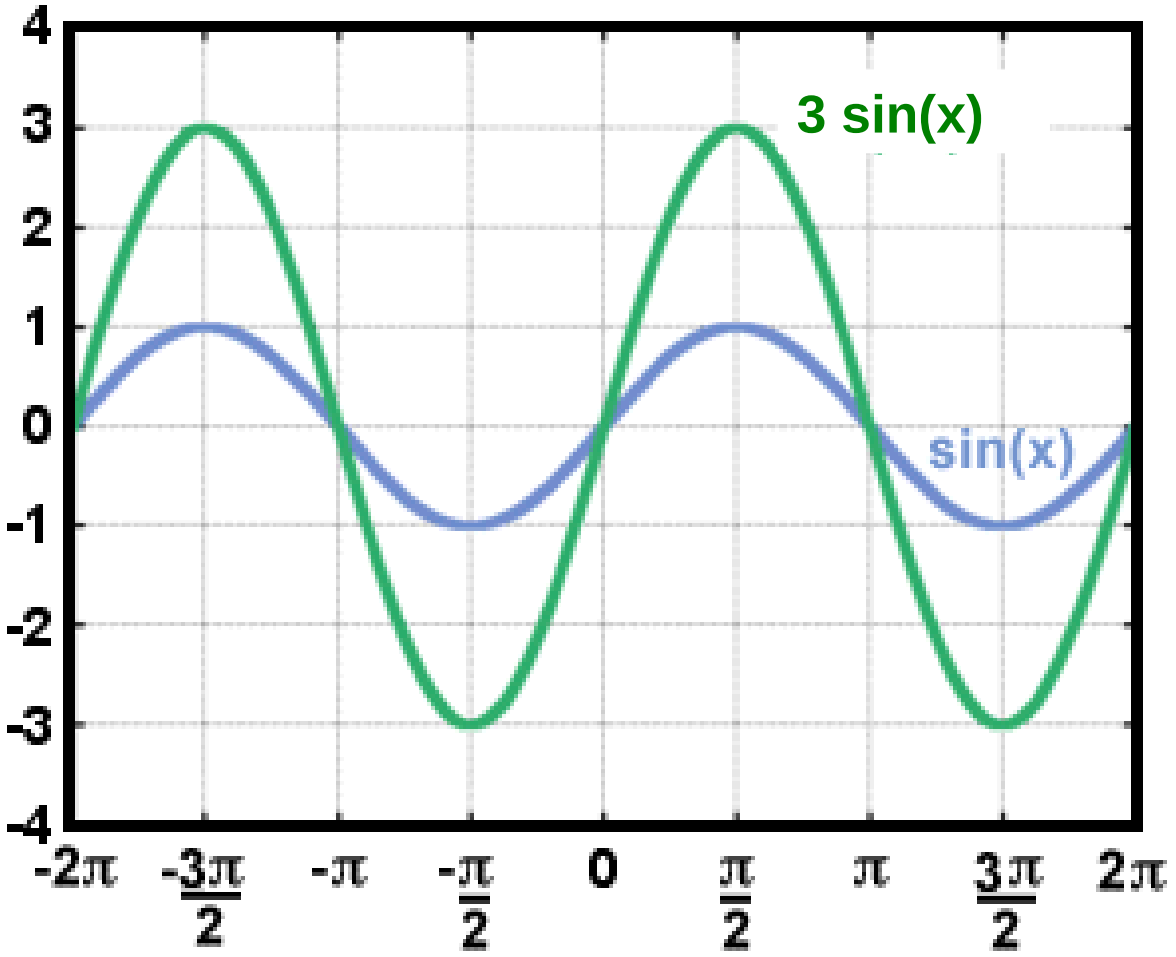
[http://en.wikipedia.org/wiki/Frequency]

Frequency vs Period

Frequency	1 mHz (10^{-3})	1 Hz (10^{0})	1 kHz (10^{3})	1 MHz (10^{6})	1 GHz (10^{9})	1 THz (10^{12})
Period (time)	1 ks (10^{3})	1 s (10^{0})	1 ms (10^{-3})	1 μs (10^{-6})	1 ns (10^{-9})	1 ps (10^{-12})

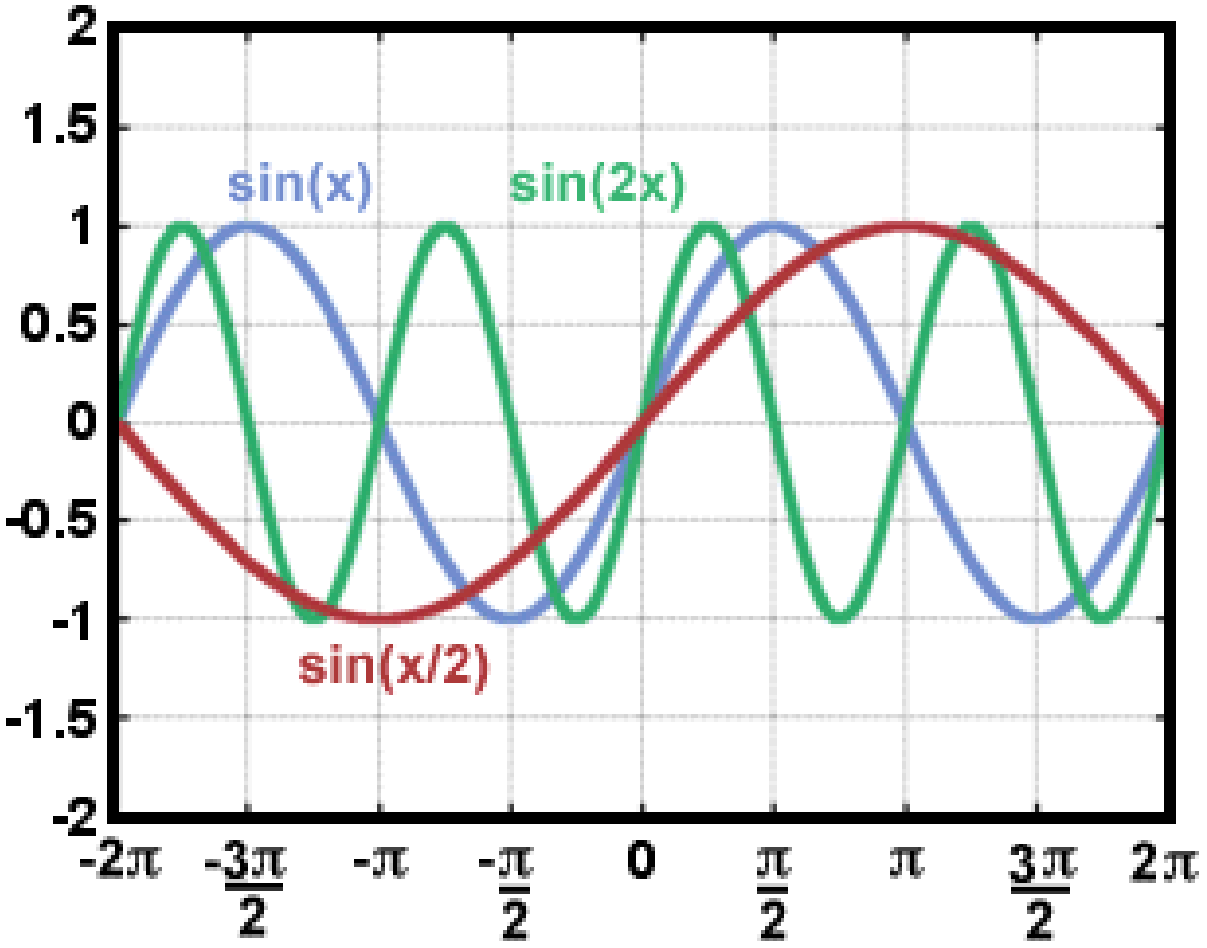
$$T = \frac{1}{f}$$

Amplitude
(vertical stretch)



[http://www.sparknotes.com/math/trigonometry/graphs/section4.rhtml]

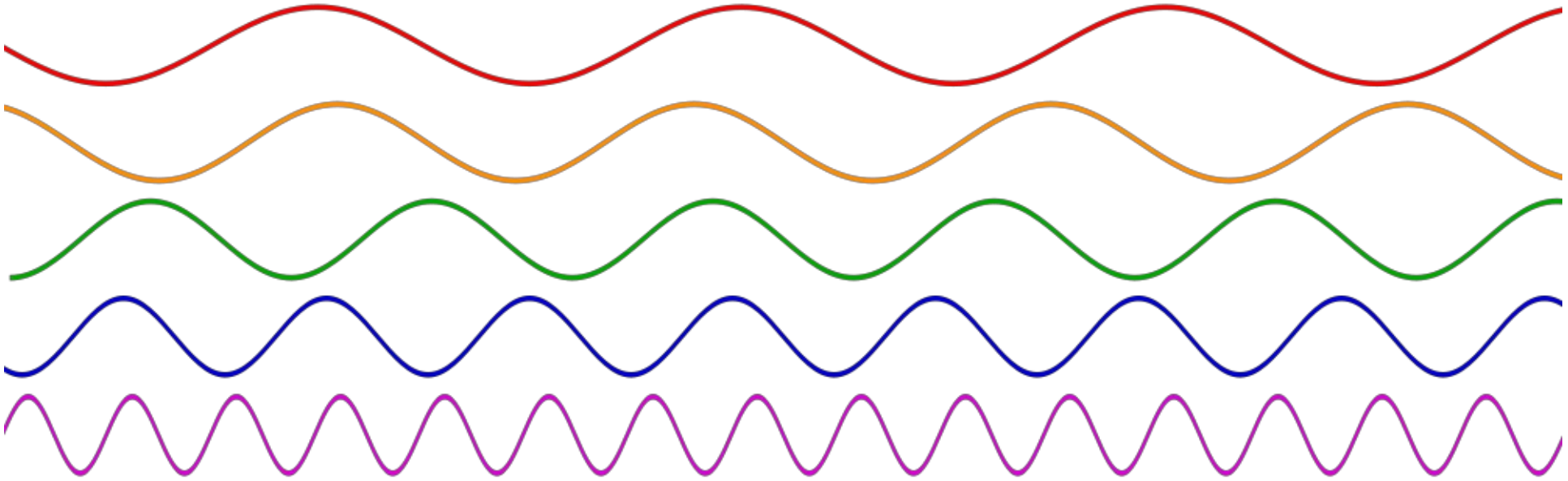
Frequency
(horizontal stretch)



[http://www.sparknotes.com/math/trigonometry/graphs/section4.rhtml]

Sinusoidal waves of various frequencies

Low Frequency

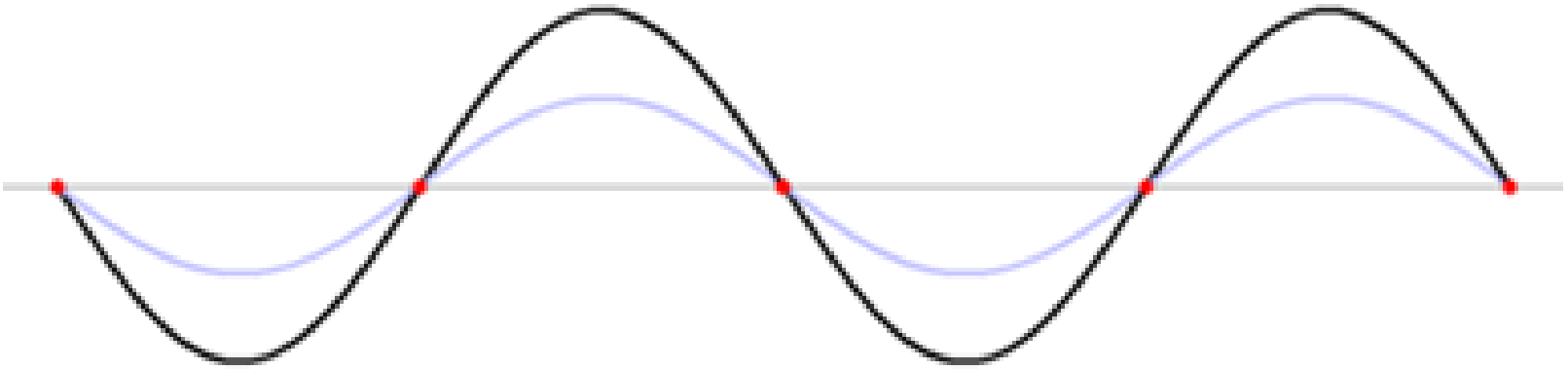


High Frequency

[http://en.wikipedia.org/wiki/Frequency]

Standing Wave

(shown in black, equal to the sum of the red and the blue waves traveling in opposite directions)

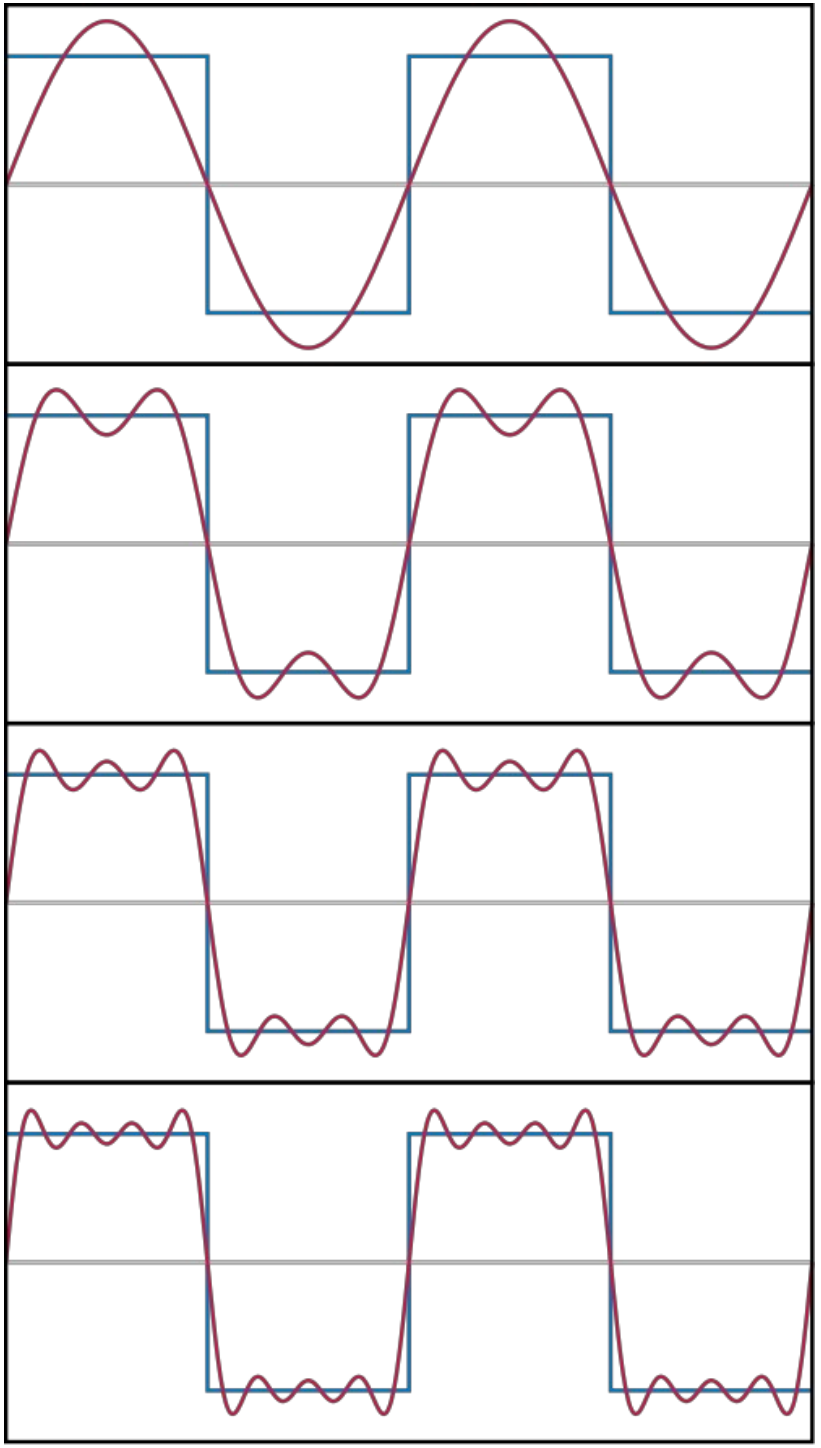


[http://en.wikipedia.org/wiki/Wavelength]

Fourier Series

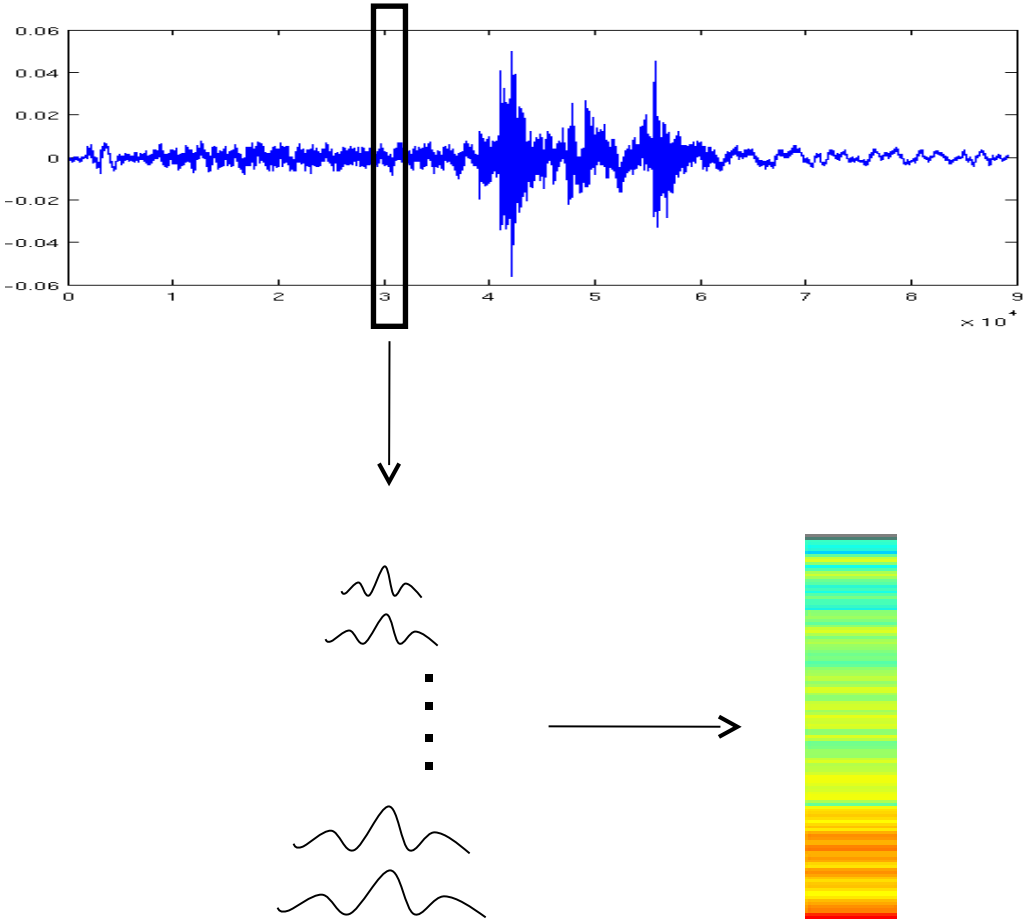
A Fourier series decomposes periodic functions or periodic signals into the sum of a (possibly infinite) set of simple oscillating functions, namely sines and cosines

Approximation

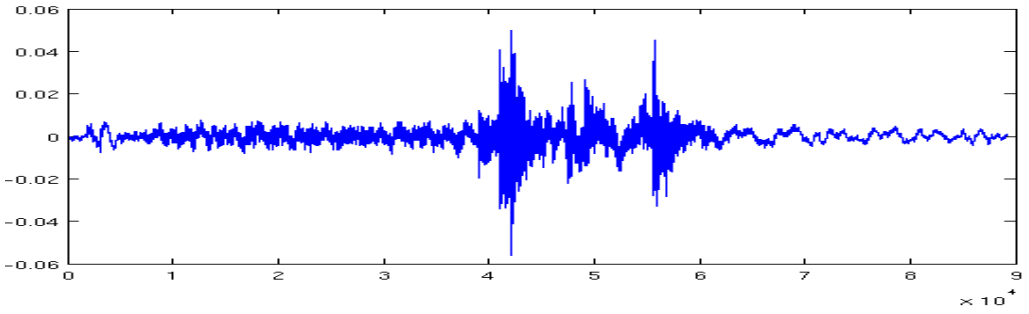


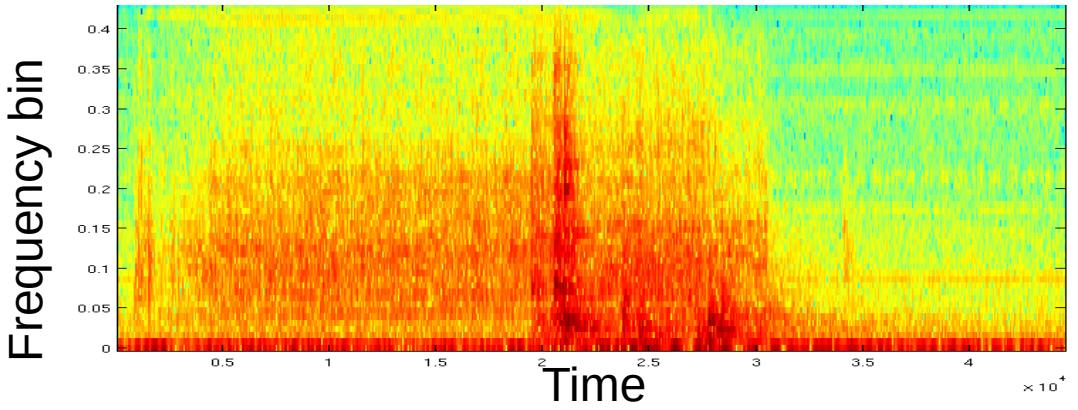
[http://en.wikipedia.org/wiki/Fourier_series]

Discrete Fourier Transform



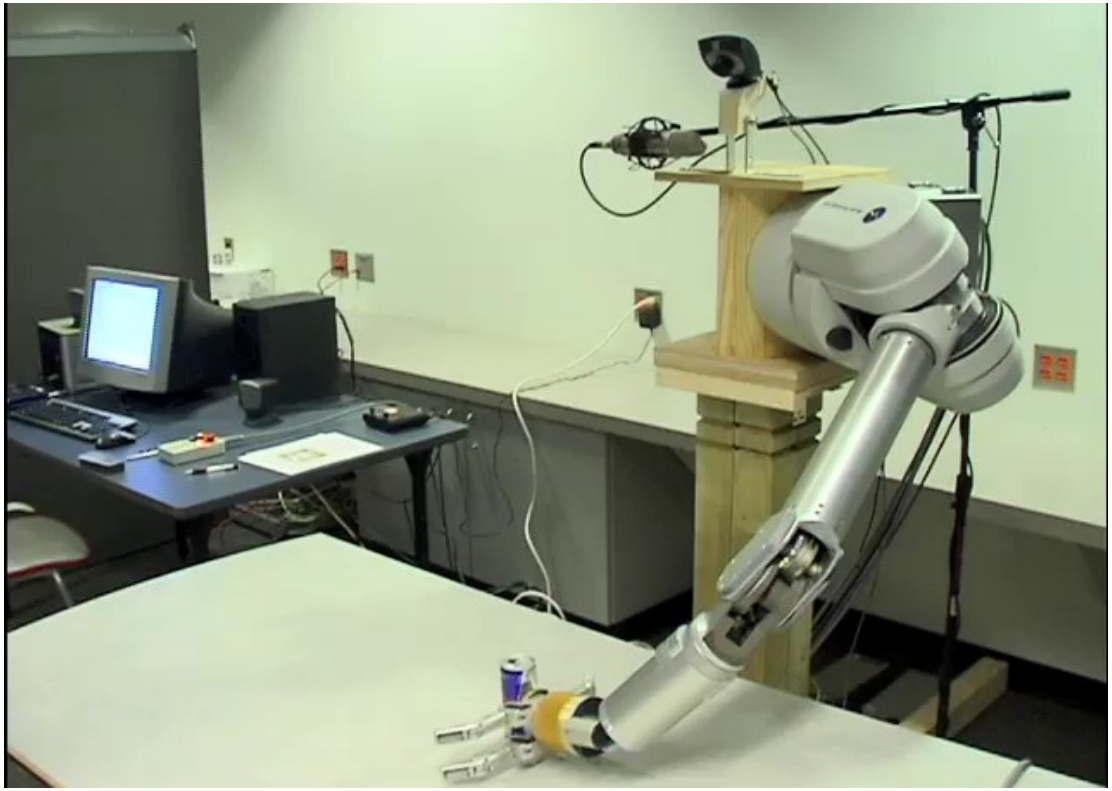
Discrete Fourier Transform





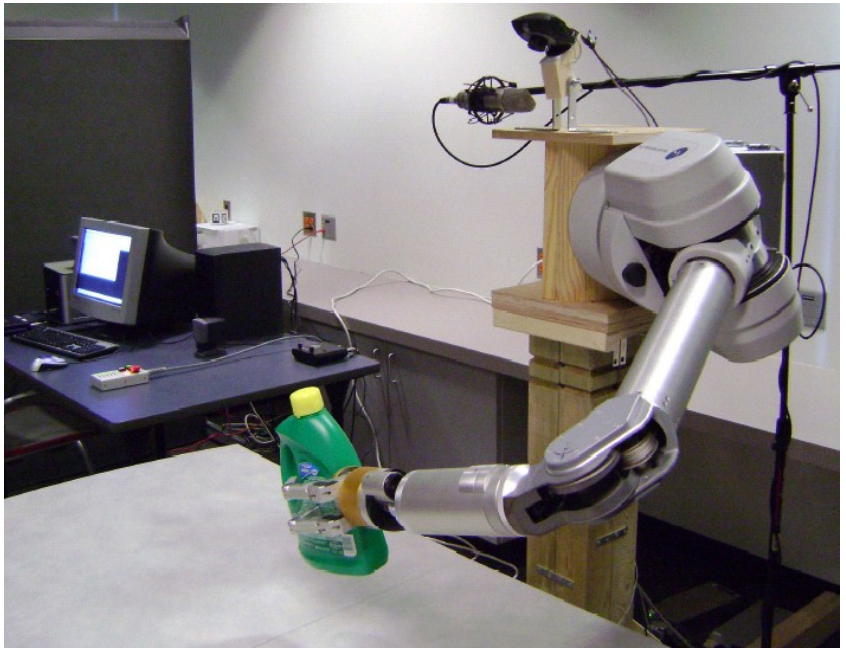
Object Exploration in Infancy

Object Exploration by a Robot



Sinapov, J., Wiemer, M., and Stoytchev, A. (2009).
Interactive Learning of the Acoustic Properties of Household Objects
In proceedings of the 2009 IEEE International Conference on Robotics and Automation (ICRA)

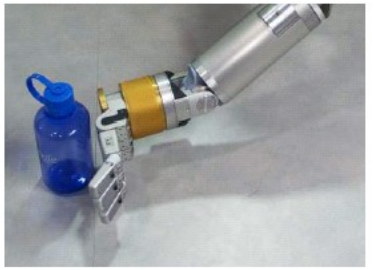
Objects



Behaviors

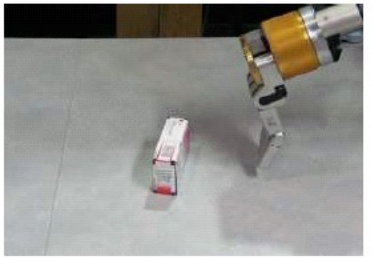
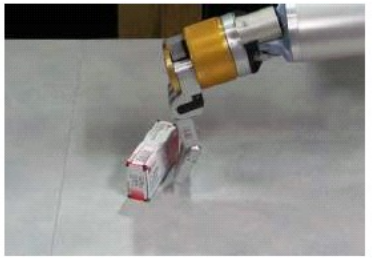
Grasp:


Push:

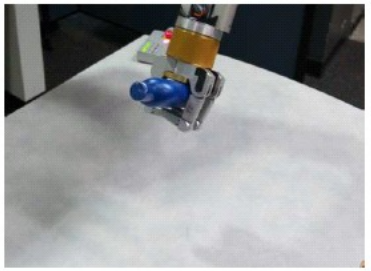
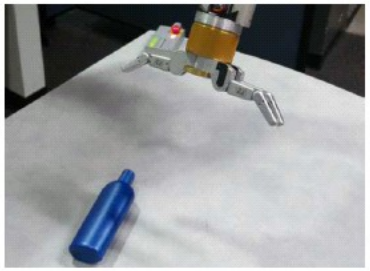



Shake:


Tap:

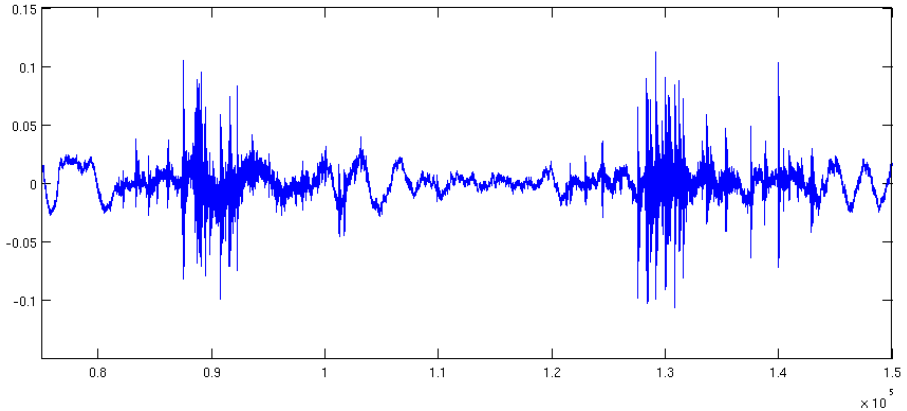
Drop:

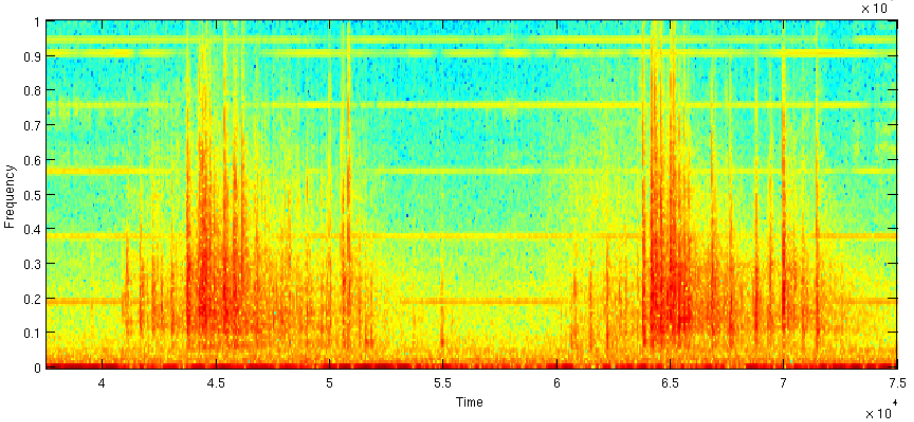
Audio Feature Extraction

Behavior Execution:

WAV file recorded:

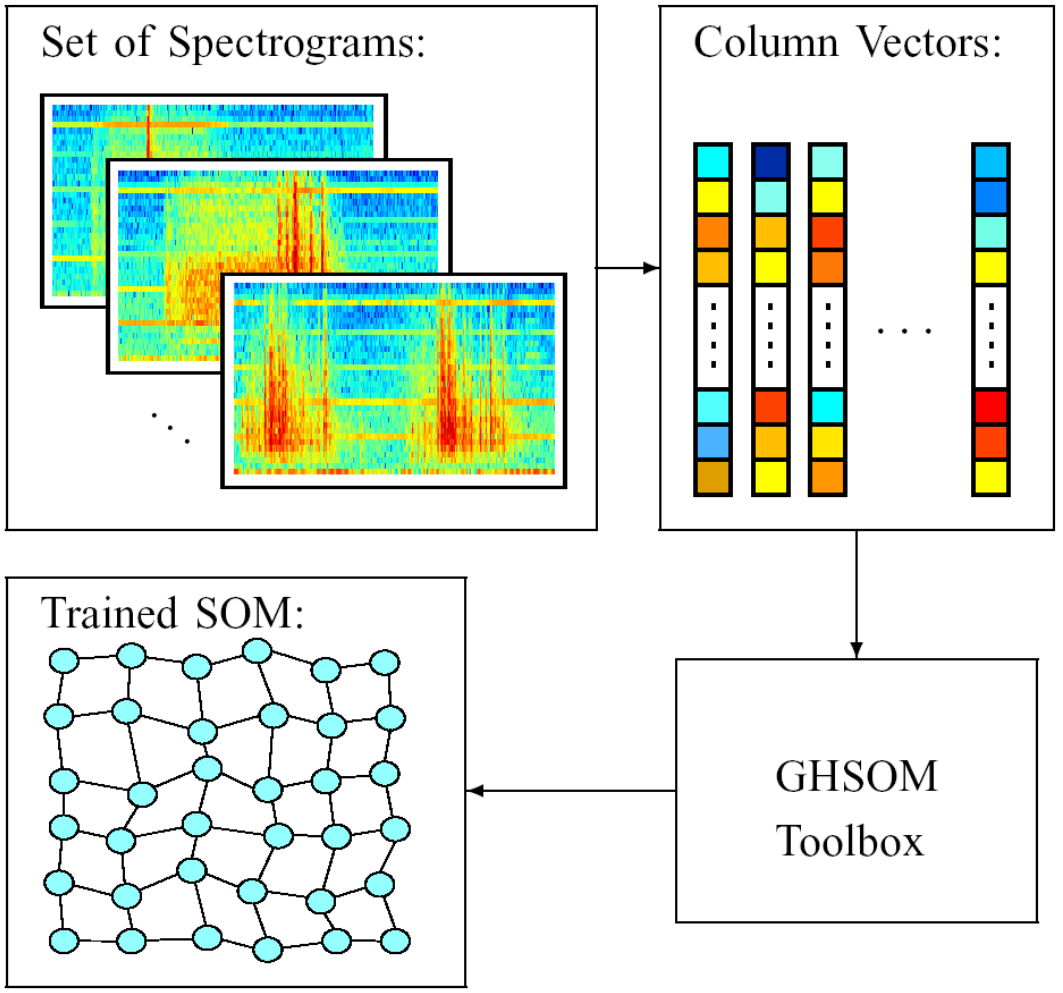


Discrete Fourier Transform:



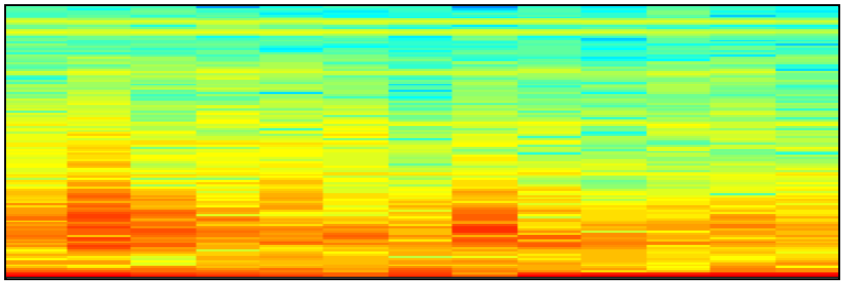
Audio Feature Extraction

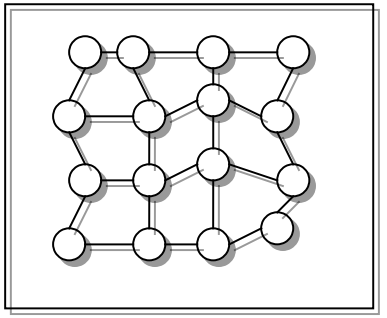
1. Training a self-organizing map (SOM) using DFT column vectors:



Audio Feature Extraction

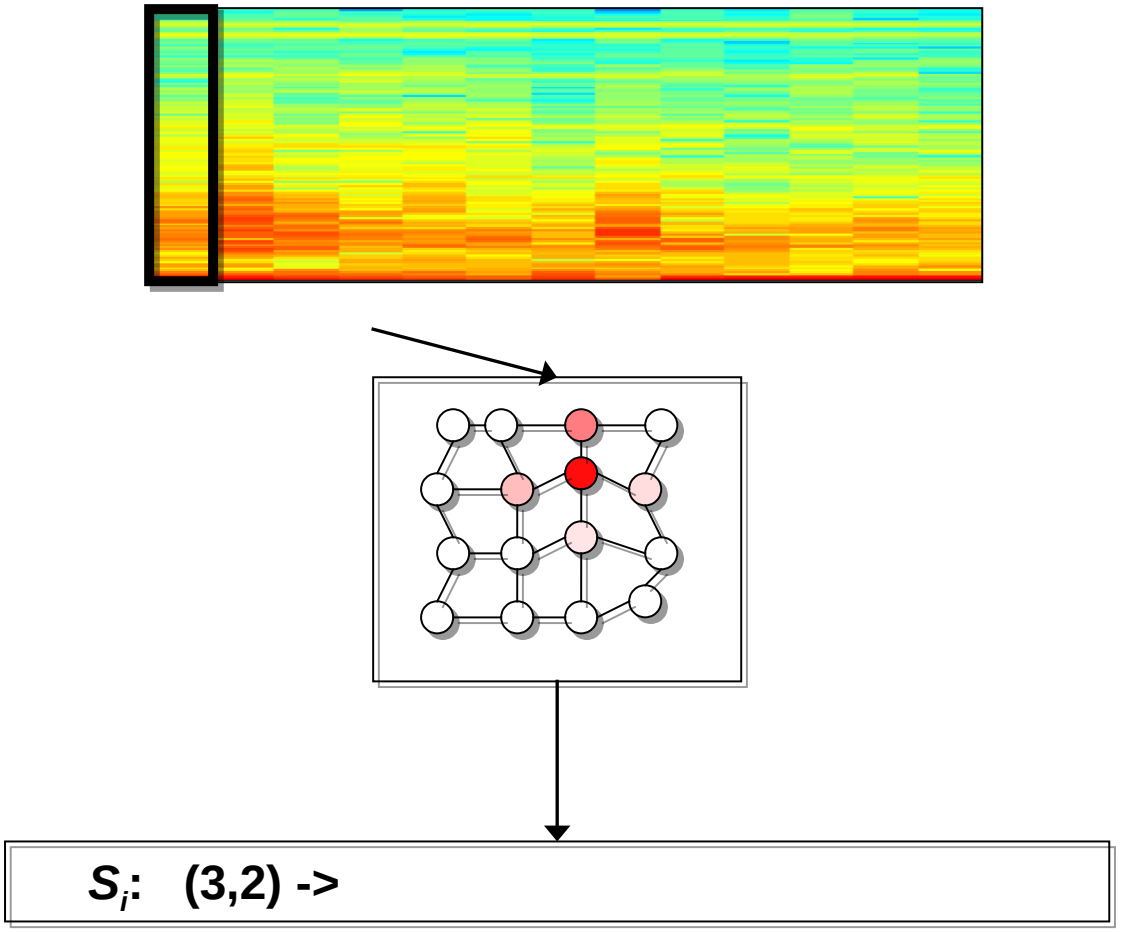
2. Use SOM to convert DFT spectrogram to a sequence:





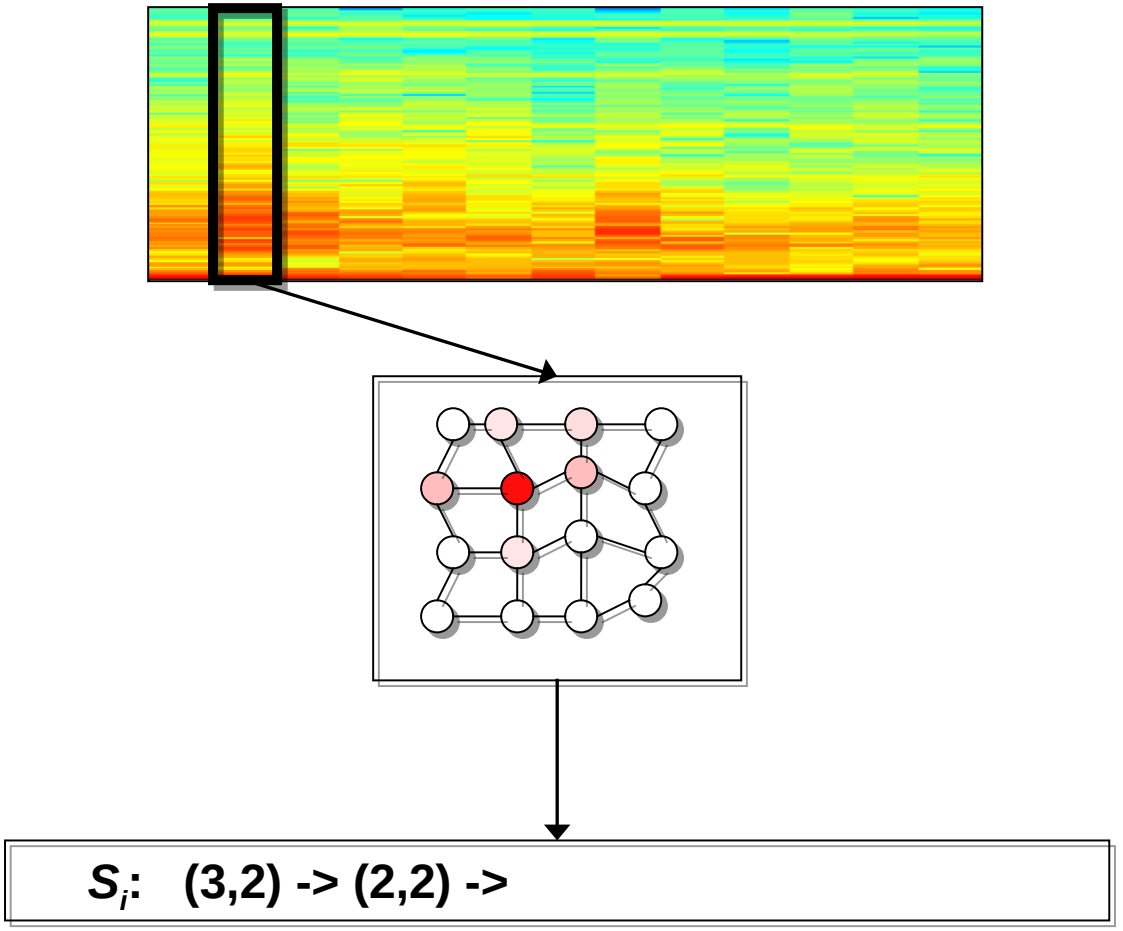
Audio Feature Extraction

2. Use SOM to convert DFT spectrogram to a sequence:



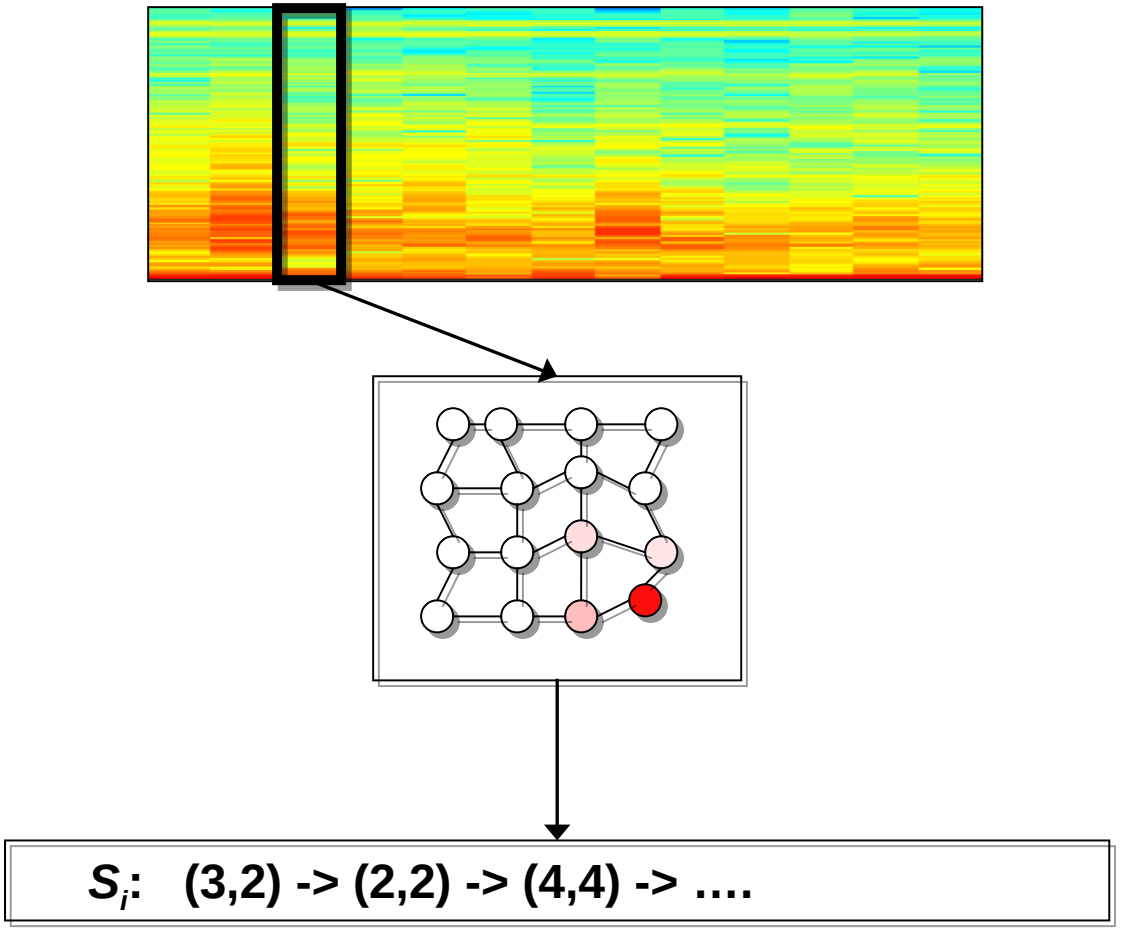
Audio Feature Extraction

2. Use SOM to convert DFT spectrogram to a sequence:



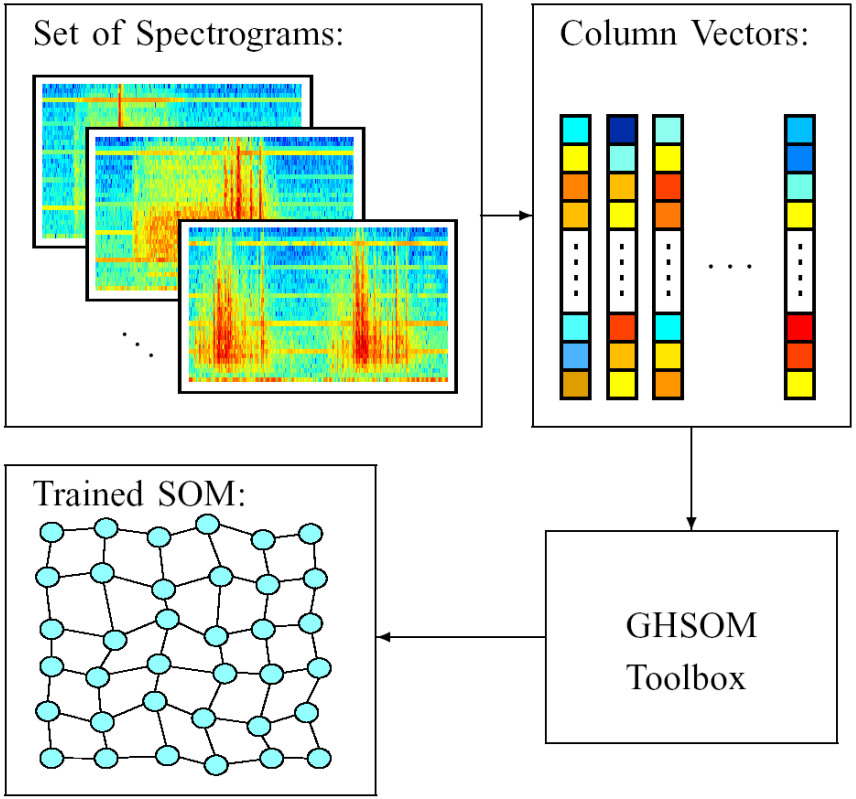
Audio Feature Extraction

2. Use SOM to convert DFT spectrogram to a sequence:

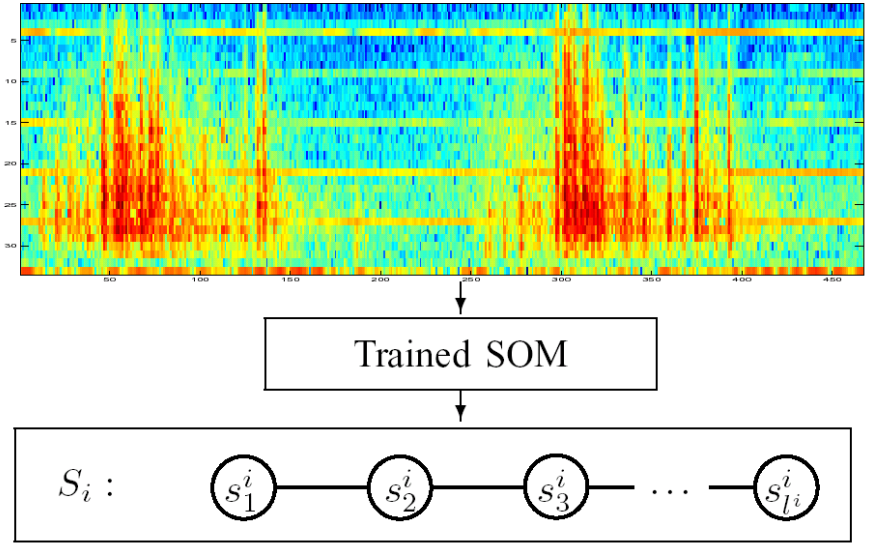


Audio Feature Extraction

1. Training a self-organizing map (SOM) using column vectors:

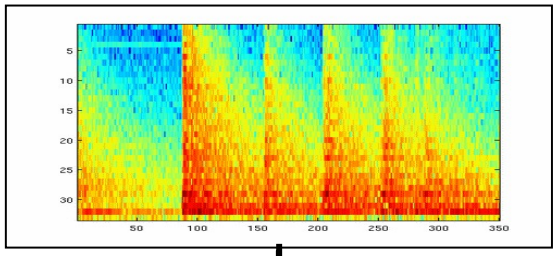


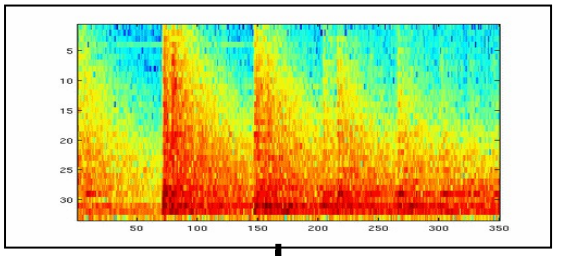
2. Discretization of a DFT of a sound using a trained SOM



$S_i = s_1^i s_2^i \dots s_{l^i}^i$ is the sequence of activated SOM nodes over the duration of the sound

Detecting Acoustic Similarity





Auditory SOM

Auditory SOM

Sequence X_i

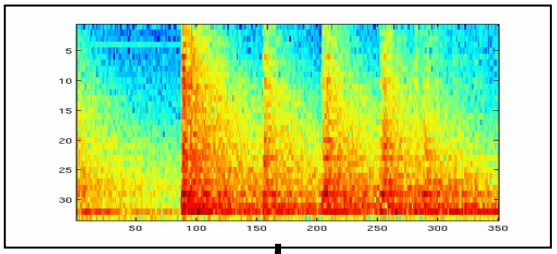
Sequence Y_j

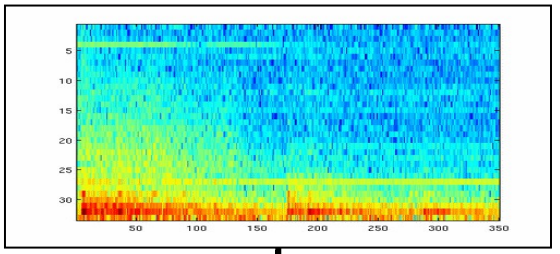
Global Sequence Alignment Similarity

$sim(X_i, Y_j) = 0.89$

very similar

Detecting Acoustic Similarity





Auditory SOM

Auditory SOM

Sequence X_i

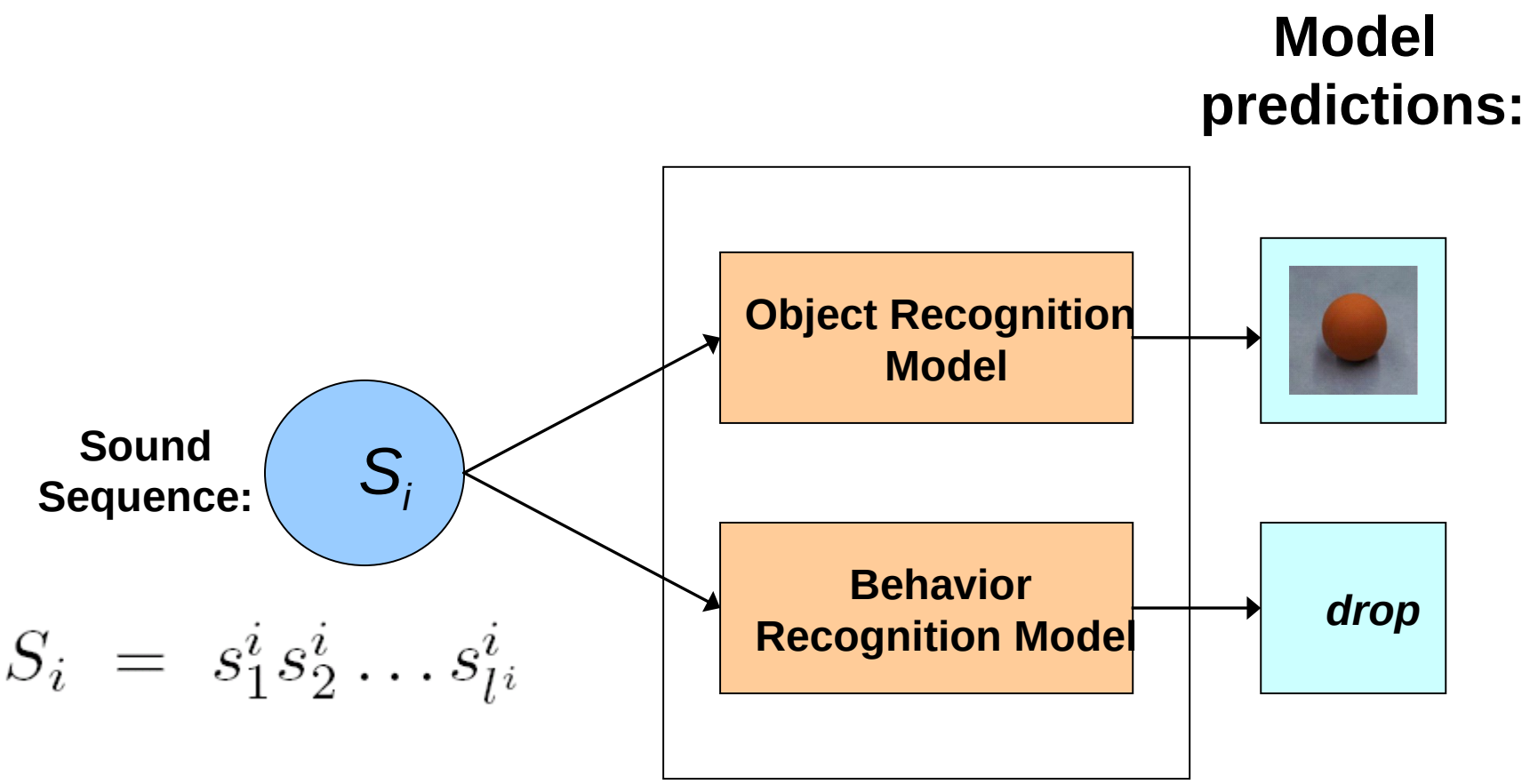
Sequence Y_j

Global Sequence Alignment Similarity

$sim(X_i,Y_j) = 0.23$

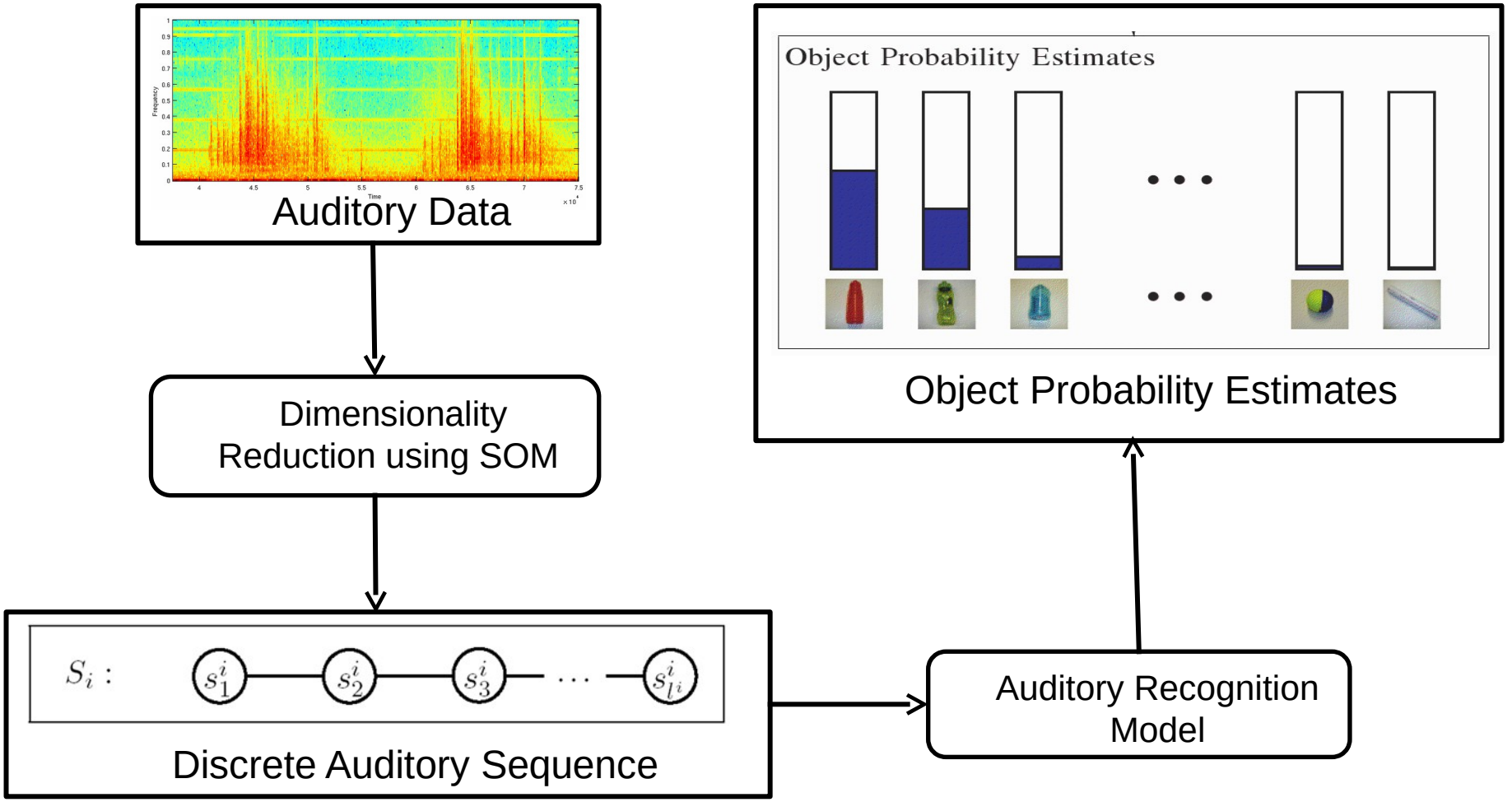
not similar

Problem Formulation



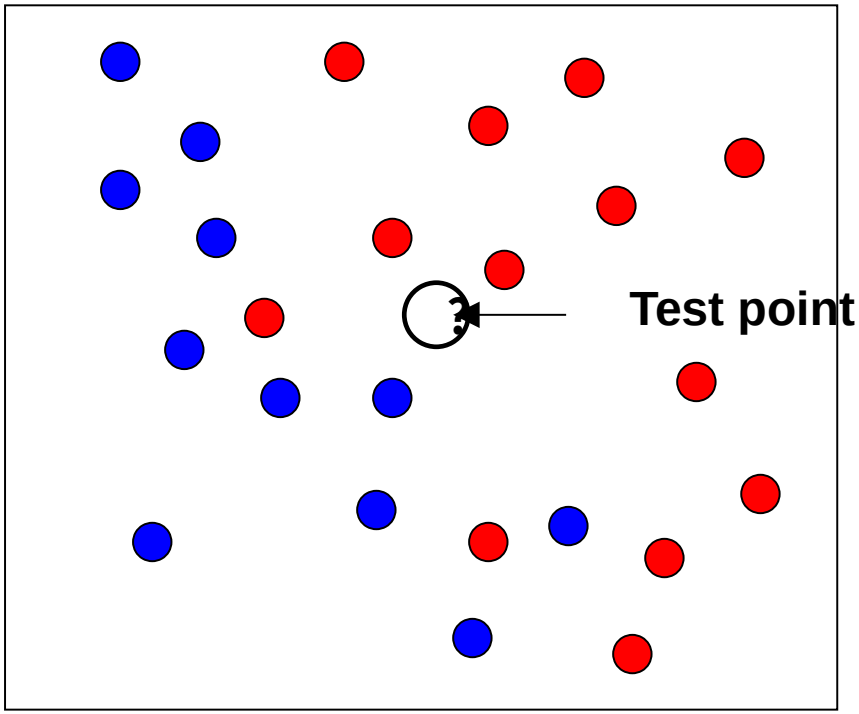
$$S_i = s_1^i s_2^i \dots s_{l^i}^i$$

Acoustic Object Recognition



Recognition Model

- k-NN: ***memory-based*** learning algorithm



With k = 3:

2 neighbors

1 neighbors

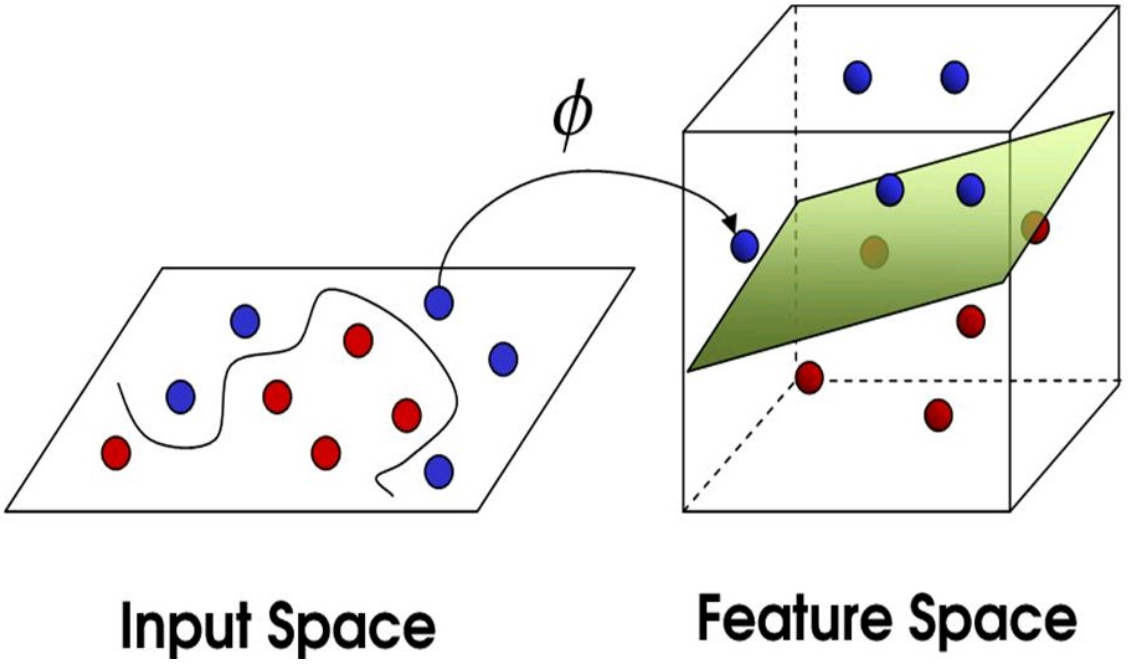
Therefore,

Pr(red) = 0.66

Pr(blue) = 0.33

Recognition Model

- SVM: ***discriminative*** learning algorithm



Evaluation Results

Behavior	k-Nearest Neighbor	Support Vector Machine
Grasp	67.89 %	75.27 %
Shake	49.47 %	50.56 %
Drop	85.79 %	80.56 %
Push	82.89 %	84.44 %
Tap	78.15 %	75.84 %
Average	72.84 %	73.33 %

Chance accuracy = 2.7 %

Evaluation Results

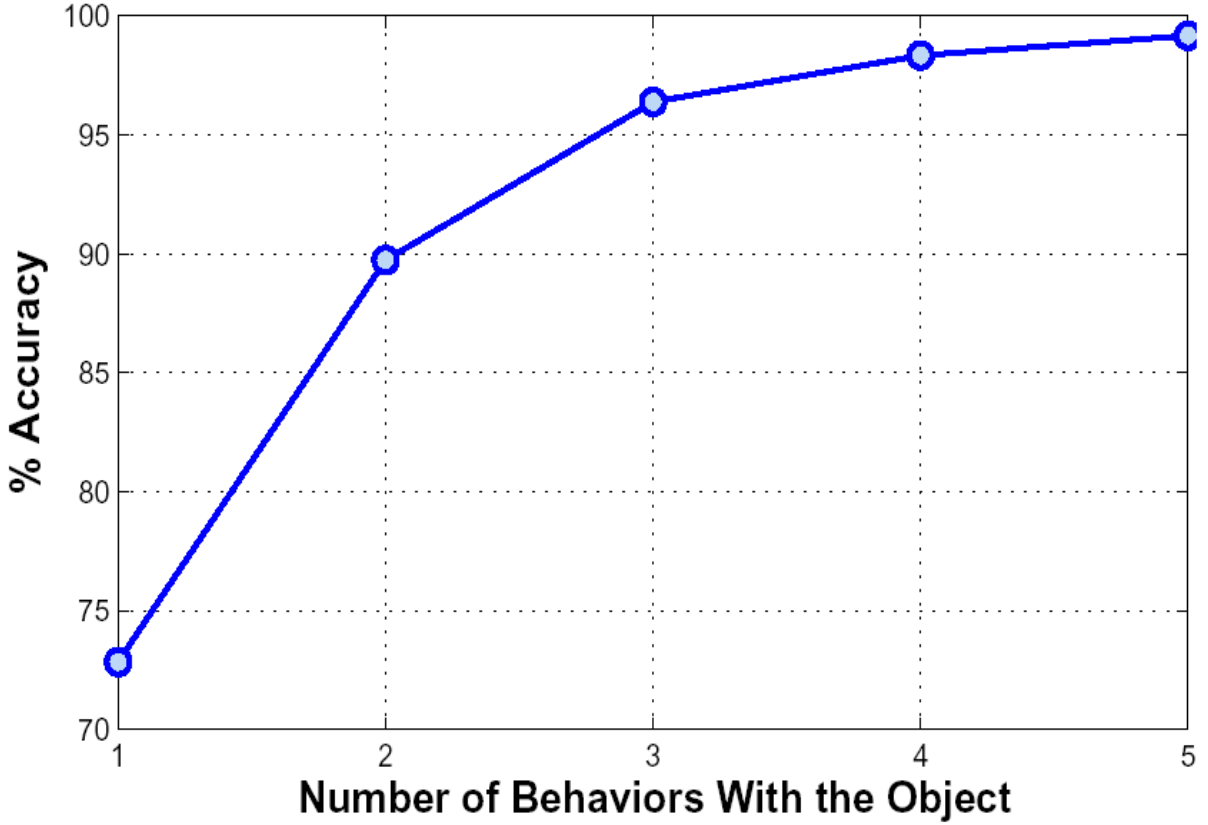
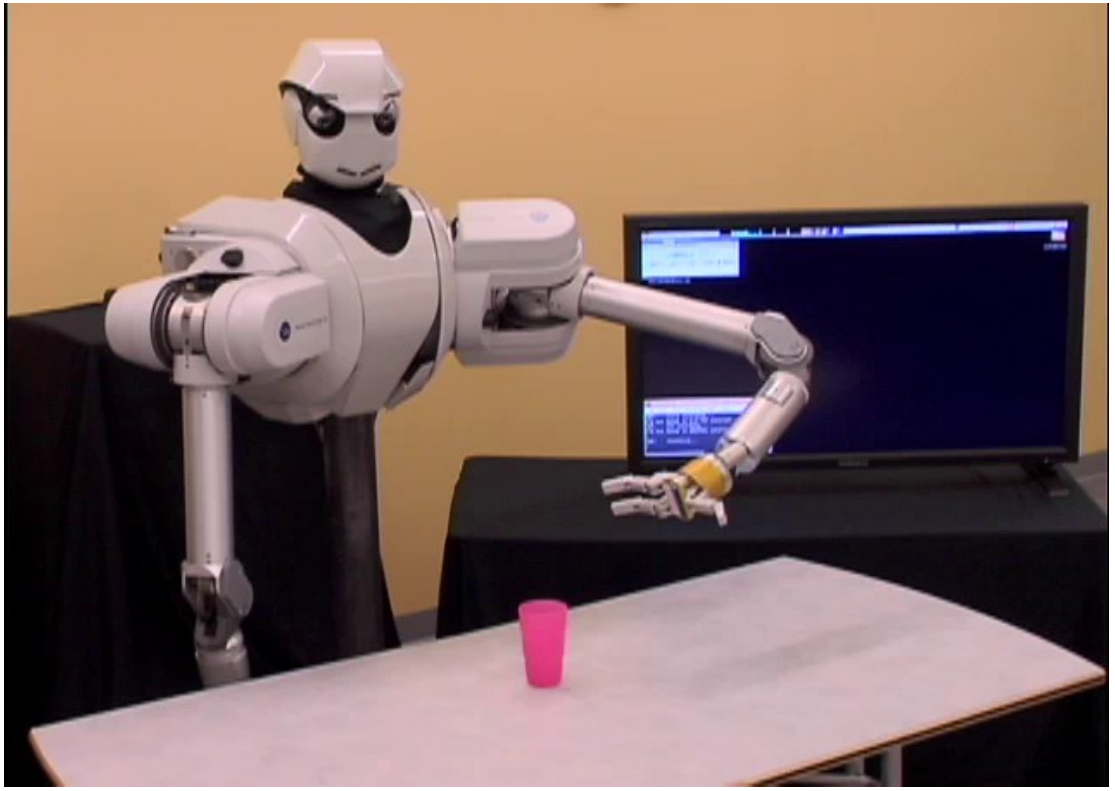
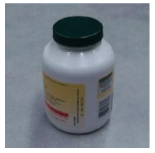


Fig. 6. Object recognition performance with k-Nearest Neighbor as the number of interactions with the object is varied from 1 (the default, used to generate Table I) to 5 (applying all five behaviors on the object).

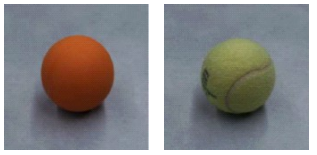
Recognition Video



Estimating Acoustic Object Similarity using Confusion Matrix

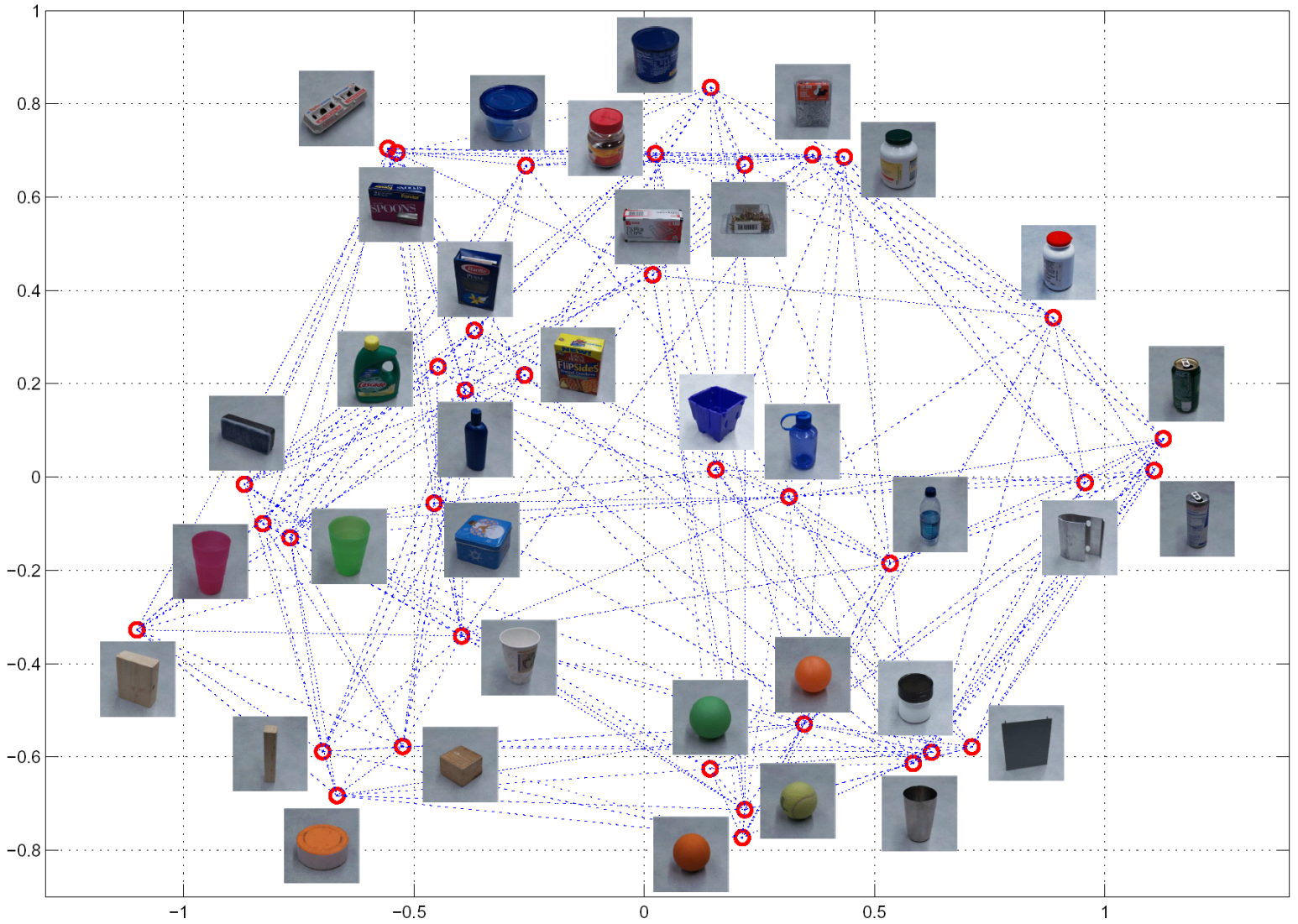
Actual \ Predicted →				
	**40**	**4**	0	0
	**6**	**42**	0	0
	0	0	**21**	**6**
	0	0	**8**	**35**

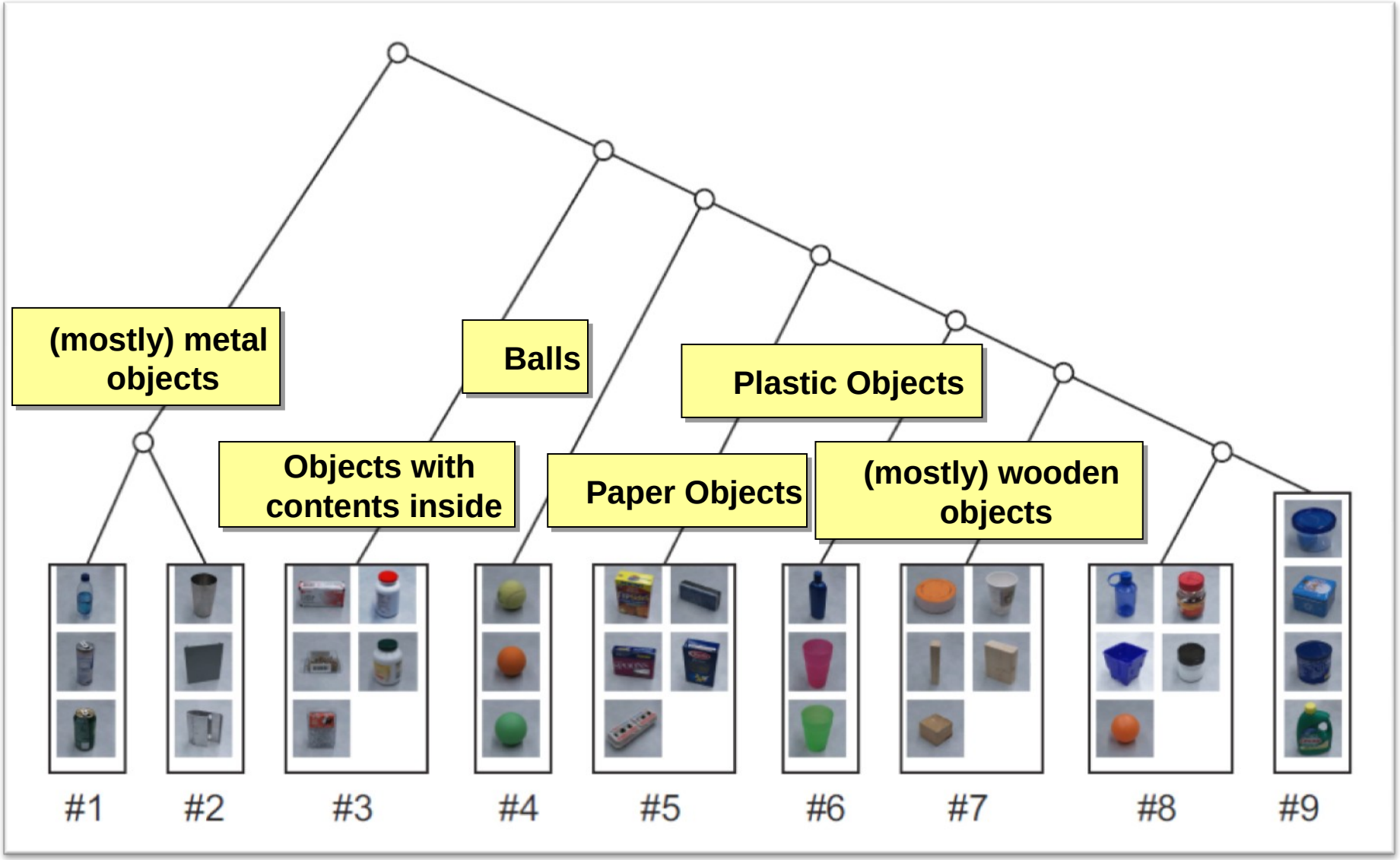

: similar


: similar

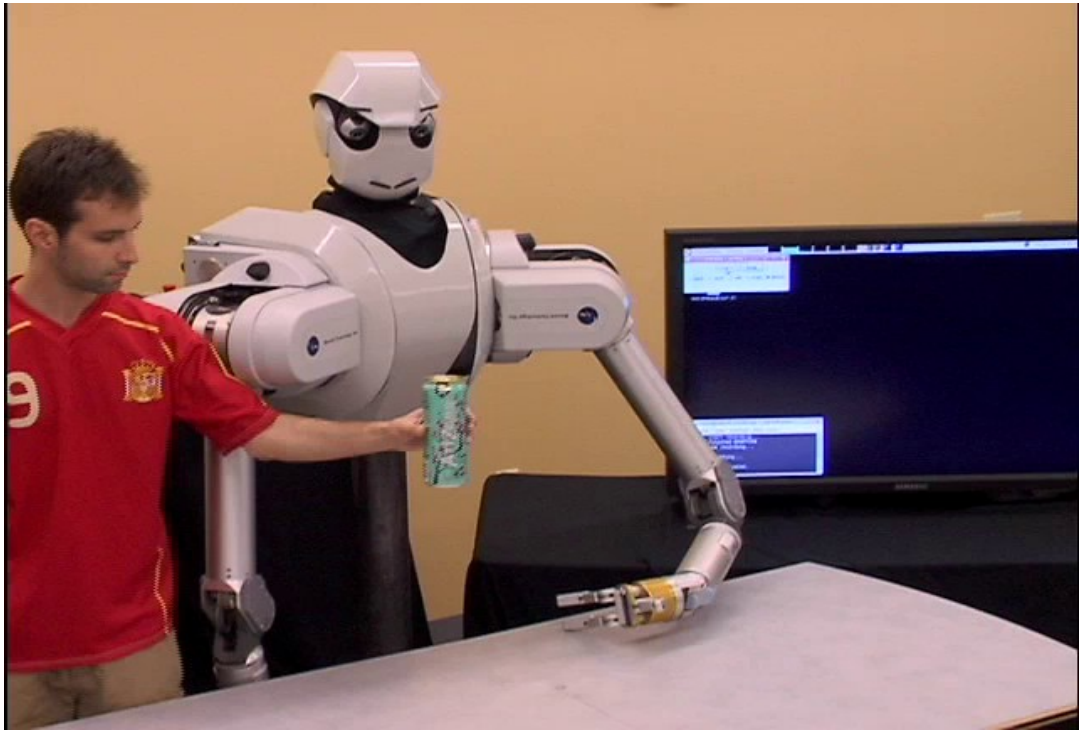

: different

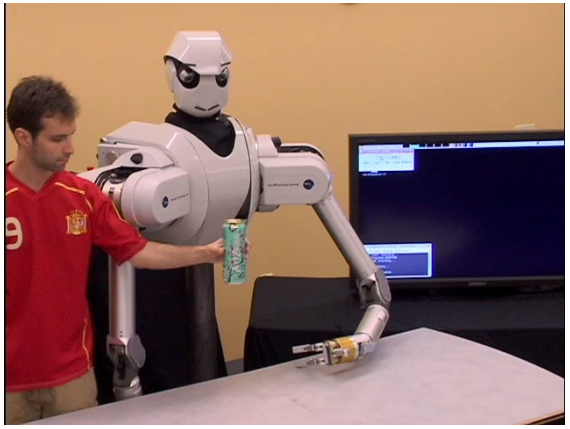

: different

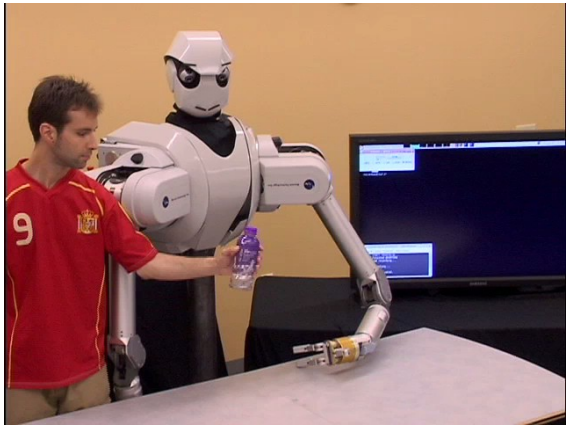

1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
−0.5
0
0.5
1

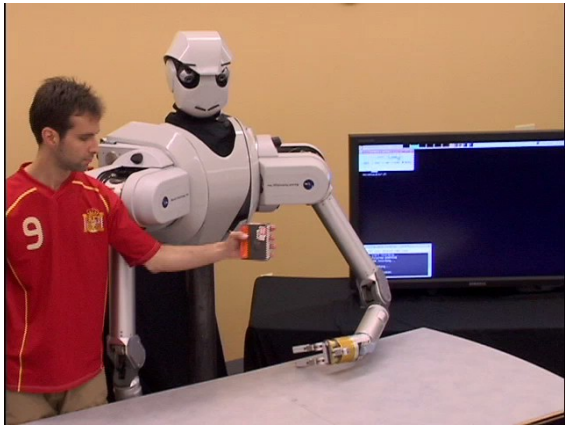

(mostly) metal objects
Balls
Plastic Objects
Objects with contents inside
Paper Objects
(mostly) wooden objects
#1
#2
#3
#4
#5
#6
#7
#8
#9

Recognizing the sounds of objects manipulated by other agents







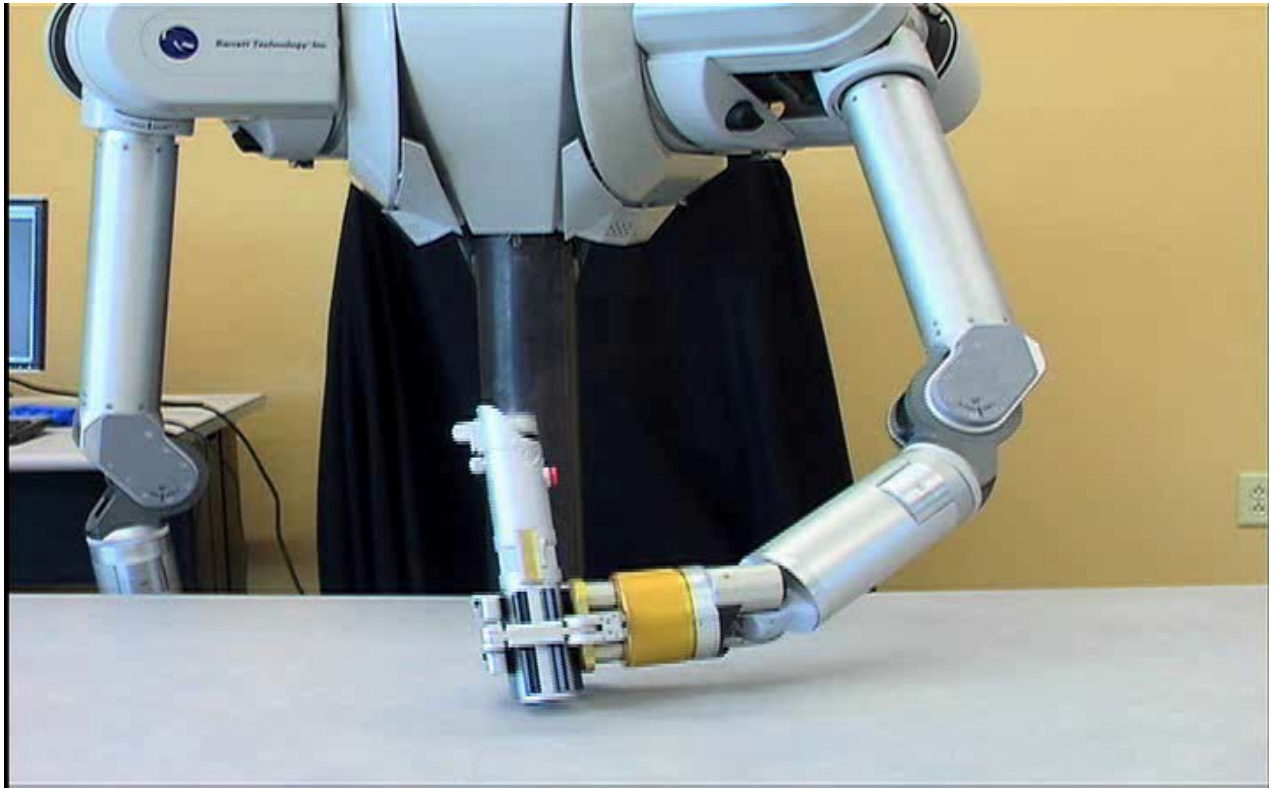


What types of sounds should a service robot operating in human environments pay attention to?

MAKING SENSE OF THE SENSES

There are many opinions about how many senses we have

SENSORY MODALITY	Conservative	Accepted	Radical
Vision	■		
Light		■	■
Colour		■	
Red			■
Green			■
Blue			■
Hearing	■	■	■
Smell	■	■	
2000 or more receptor types			■
Taste	■		
Sweet		■	■
Salt		■	■
Sour		■	■
Bitter		■	■
Umami			■
Touch	■	■	
Light touch			■
Pressure			■
Pain	■	■	
Cutaneous			■
Somatic			■
Visceral			■
Mechanoreception	■		
Balance	■	■	
Rotational acceleration			■
Linear acceleration			■
Proprioception – joint position		■	■
Kinaesthesis		■	
Muscle stretch – Golgi tendon organs			■
Muscle stretch – muscle spindles			■
Temperature	■		
Heat		■	■
Cold		■	■
Interoceptors			
Blood pressure	■	■	
Arterial blood pressure			■
Central venous blood pressure			■
Head blood temperature			■
Blood oxygen content		■	■
Cerebrospinal fluid pH		■	■
Plasma osmotic pressure (thirst?)		■	■
Artery-vein blood glucose difference (hunger?)		■	■
Lung inflation		■	■
Bladder stretch			■
Full stomach			■
TOTAL	10	21	33

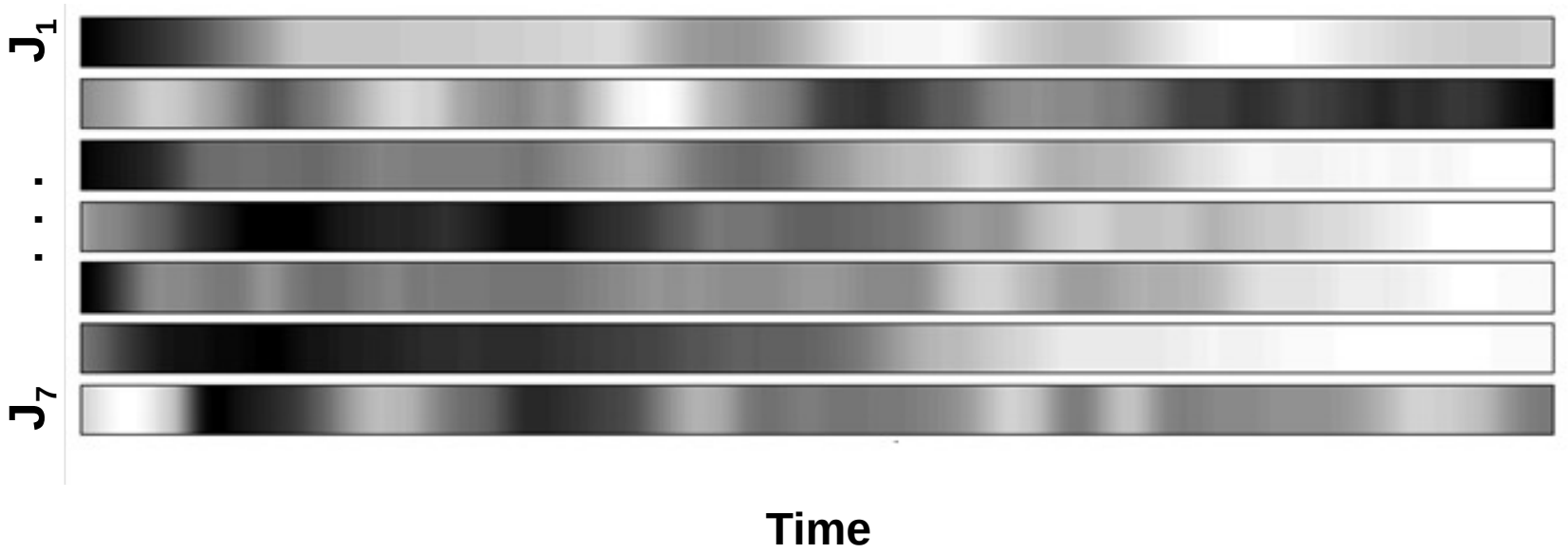


Sinapov, J., Bergquist, T., Schenck, C., Ohiri, U., Griffith, S., and Stoytchev, A. (2011)
Interactive Object Recognition Using Proprioceptive and Auditory Feedback
International Journal of Robotics Research, Vol. 30, No. 10, pp. 1250-1262, September 2011

https://youtu.be/MbLIbAL1edw

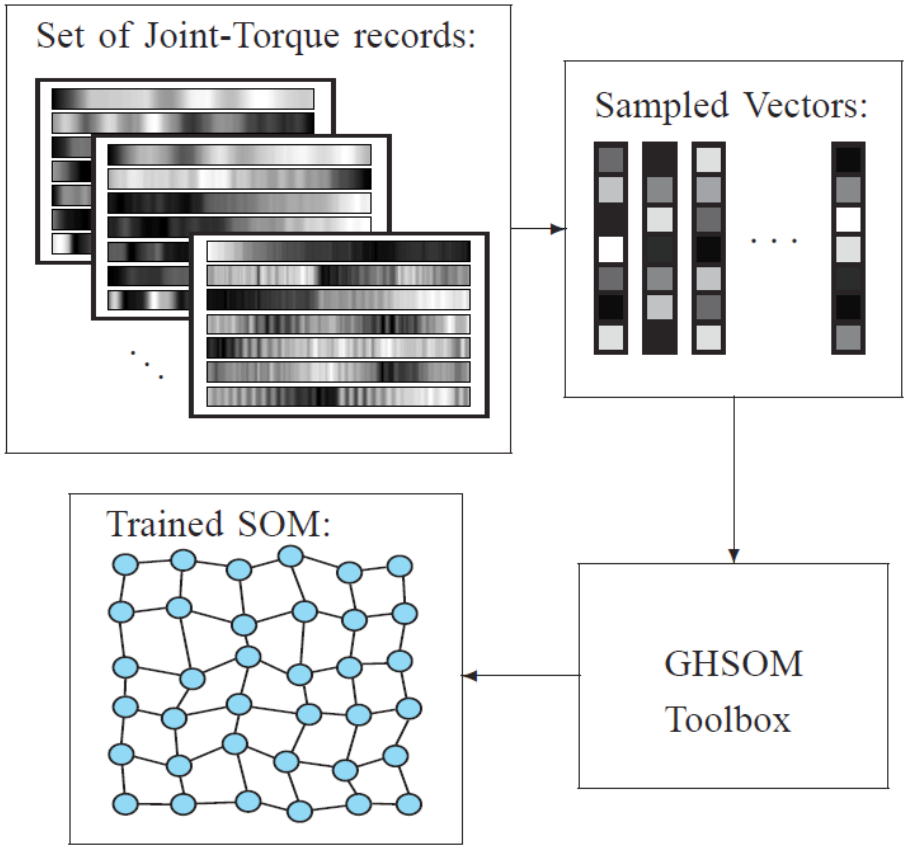
Objects

The Proprioceptive / Haptic Modality

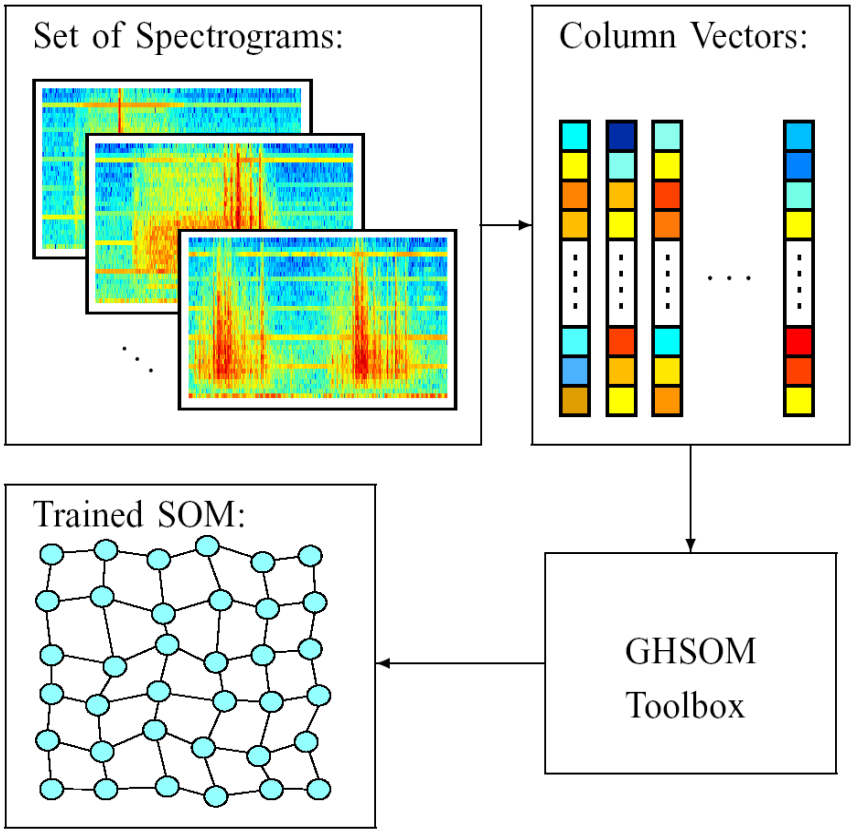


Feature Extraction

Training a self-organizing map (SOM) using sampled joint torques:

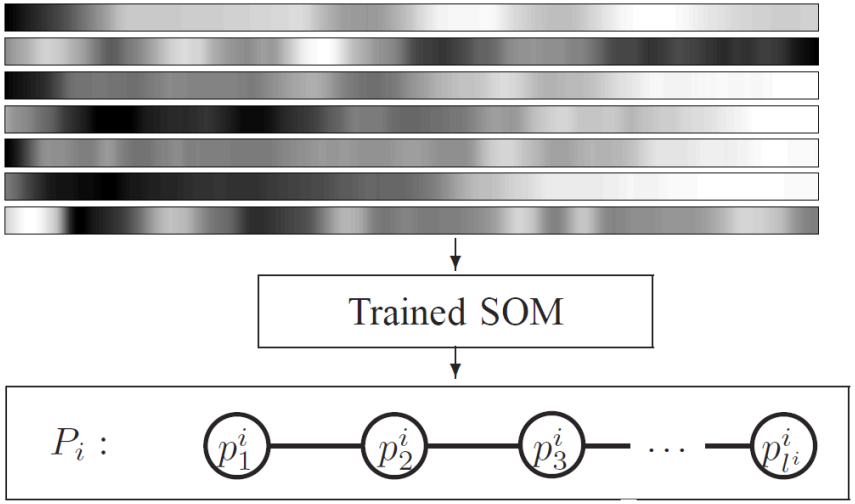


Training an SOM using sampled frequency distributions:

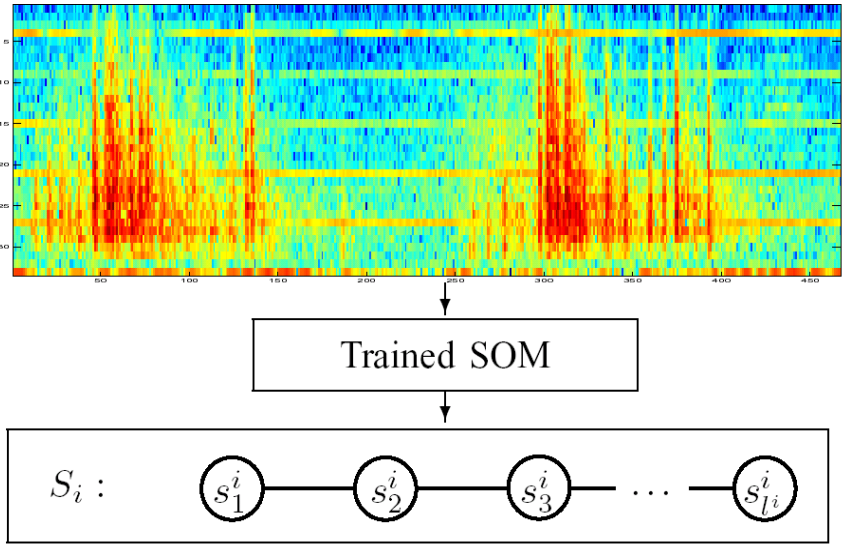


Feature Extraction

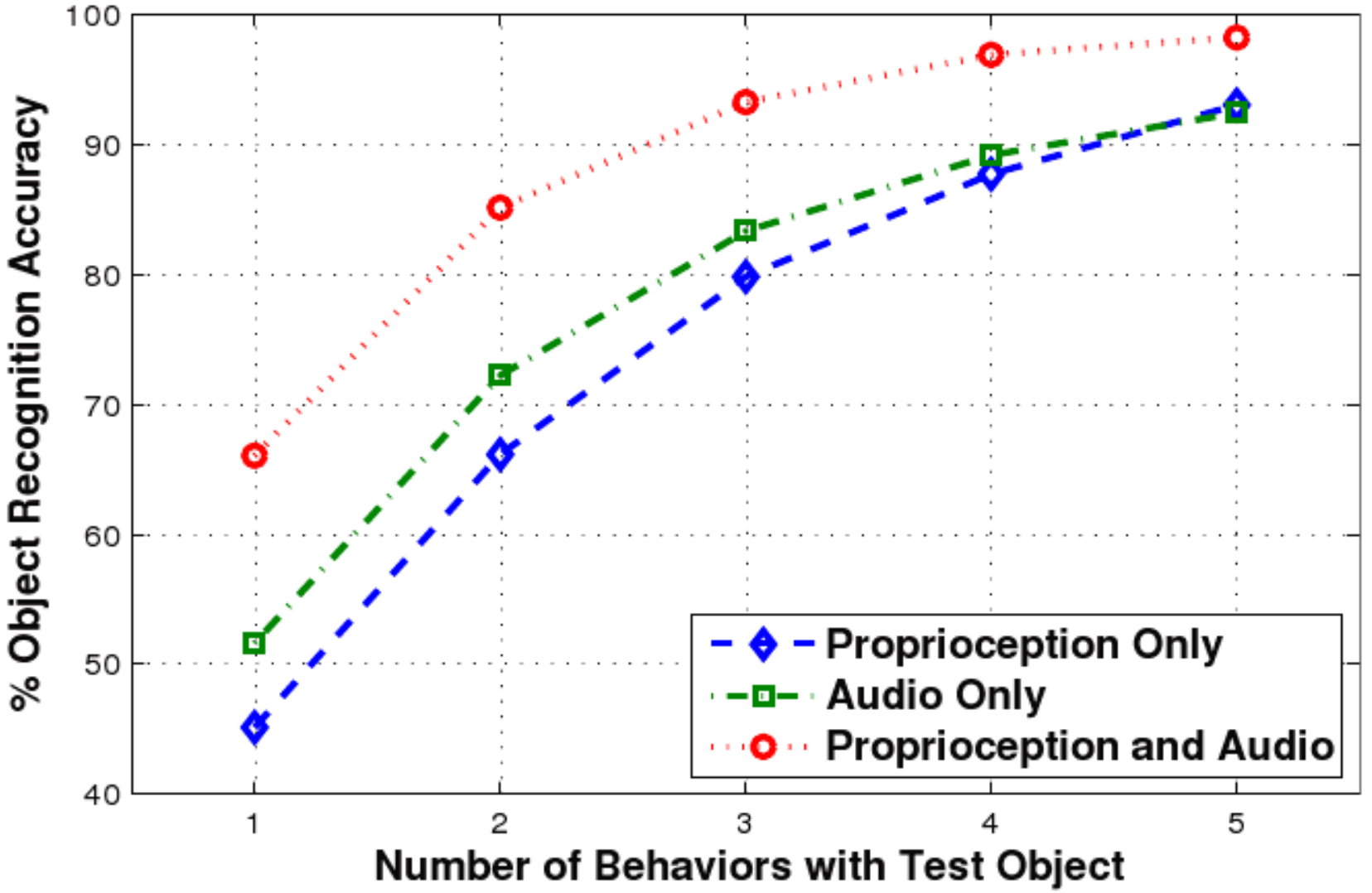
Discretization of joint-torque records using a trained SOM

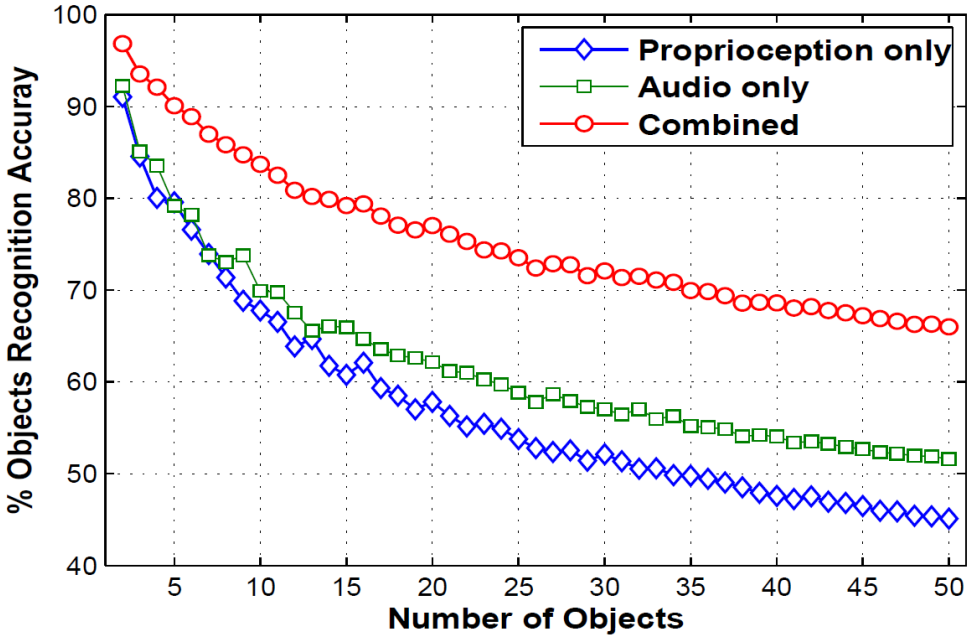


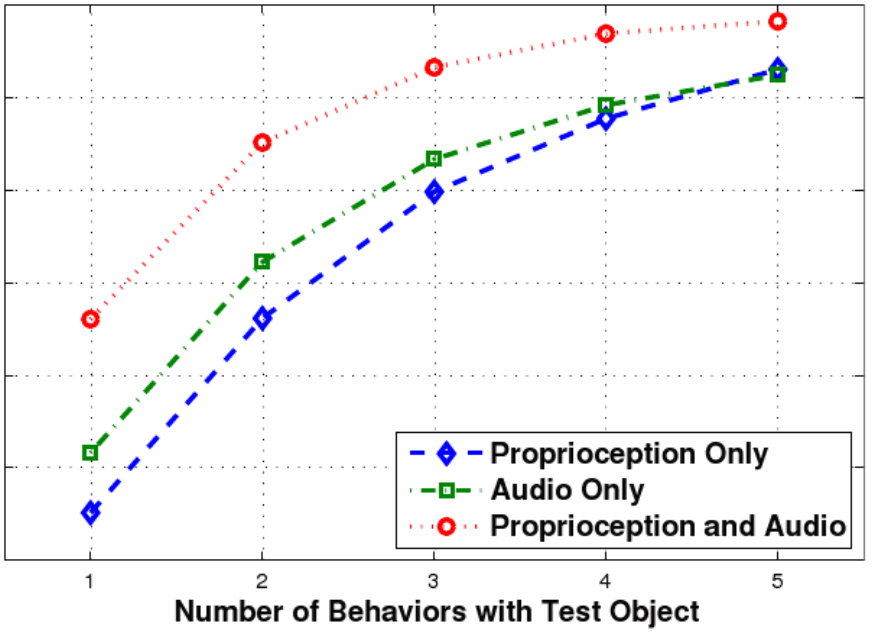
Discretization of the DFT of a sound using a trained SOM



Accuracy vs. Number of Behaviors

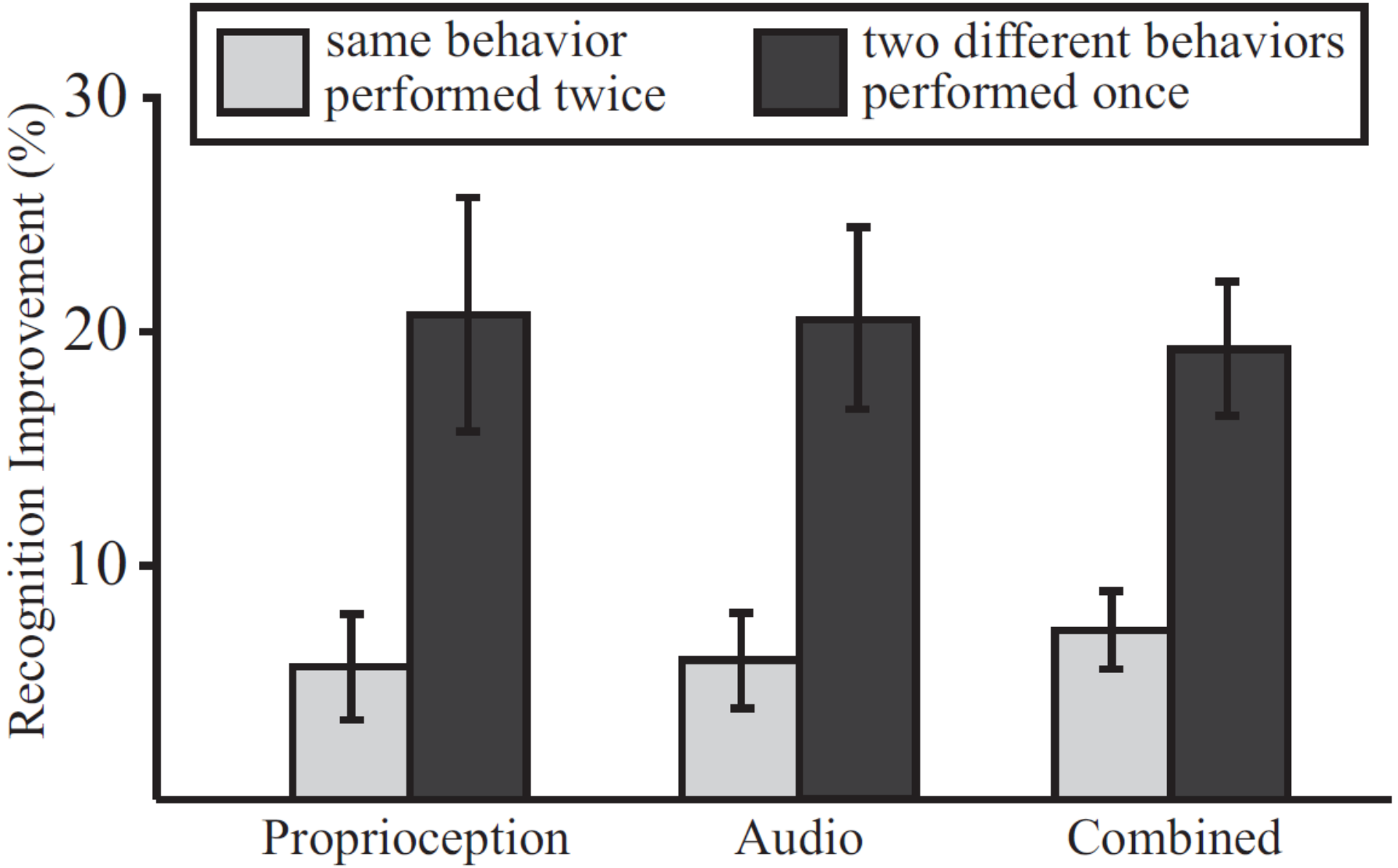






1 Behavior

Multiple Behaviors

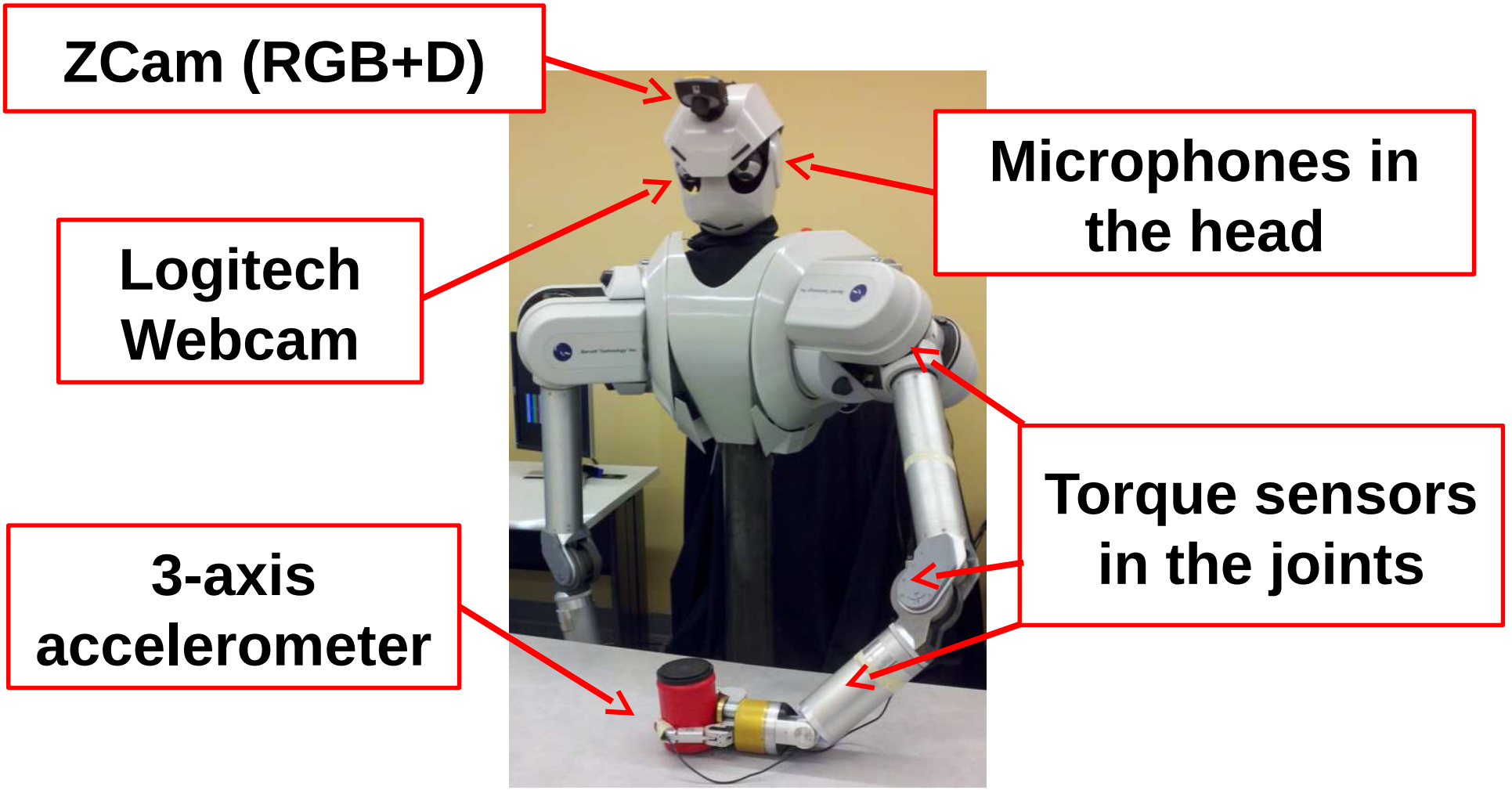


Sinapov, J., and Stoytchev, A. (2010).
The Boosting Effect of Exploratory Behaviors
In Proceedings of the 24-th National Conference on Artificial Intelligence (AAAI), 2010.

MAKING SENSE OF THE SENSES

There are many opinions about how many senses we have

SENSORY MODALITY	Conservative	Accepted	Radical
Vision	■		
Light		■	■
Colour		■	
Red			■
Green			■
Blue			■
Hearing	■	■	■
Smell	■	■	
2000 or more receptor types			■
Taste	■		
Sweet		■	■
Salt		■	■
Sour		■	■
Bitter		■	■
Umami			■
Touch	■	■	
Light touch			■
Pressure			■
Pain	■	■	
Cutaneous			■
Somatic			■
Visceral			■
Mechanoreception	■		
Balance	■	■	
Rotational acceleration			■
Linear acceleration			■
Proprioception – joint position		■	■
Kinaesthesis		■	
Muscle stretch – Golgi tendon organs			■
Muscle stretch – muscle spindles			■
Temperature	■		
Heat		■	■
Cold		■	■
Interoceptors			
Blood pressure	■	■	
Arterial blood pressure			■
Central venous blood pressure			■
Head blood temperature			■
Blood oxygen content		■	■
Cerebrospinal fluid pH		■	■
Plasma osmotic pressure (thirst?)		■	■
Artery-vein blood glucose difference (hunger?)		■	■
Lung inflation		■	■
Bladder stretch			■
Full stomach			■
TOTAL	10	21	33



Sinapov, J., Schenck, C., Staley, K., Sukhoy, V., and Stoytchev, A. (2014)
Grounding Semantic Categories in Behavioral Interactions: Experiments with 100 Objects
Robotics and Autonomous Systems, Vol. 62, No. 5, pp. 632-645, May 2014.

Exploratory Behaviors

grasp

lift

hold

shake

drop

tap

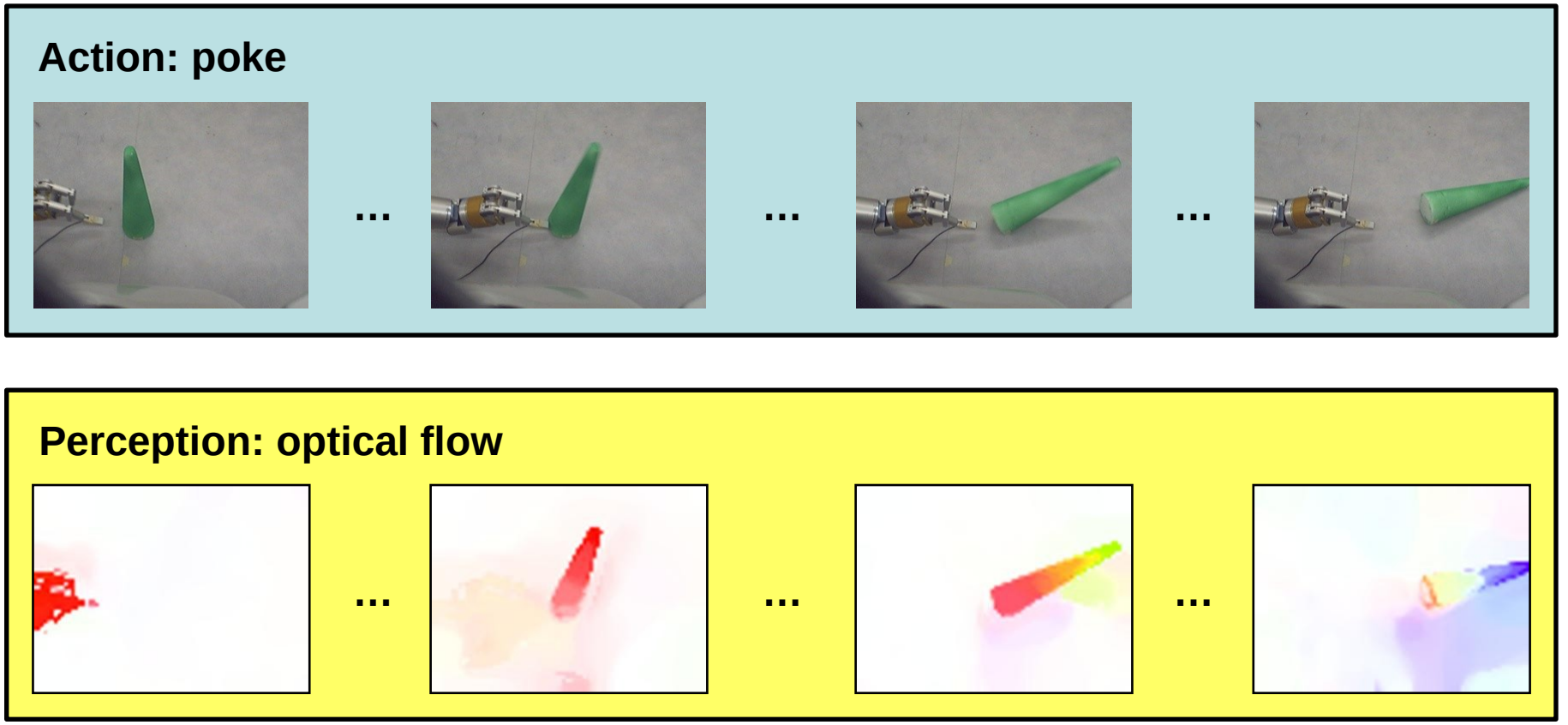
poke

push

press

https://youtu.be/FML4BuX7P0k

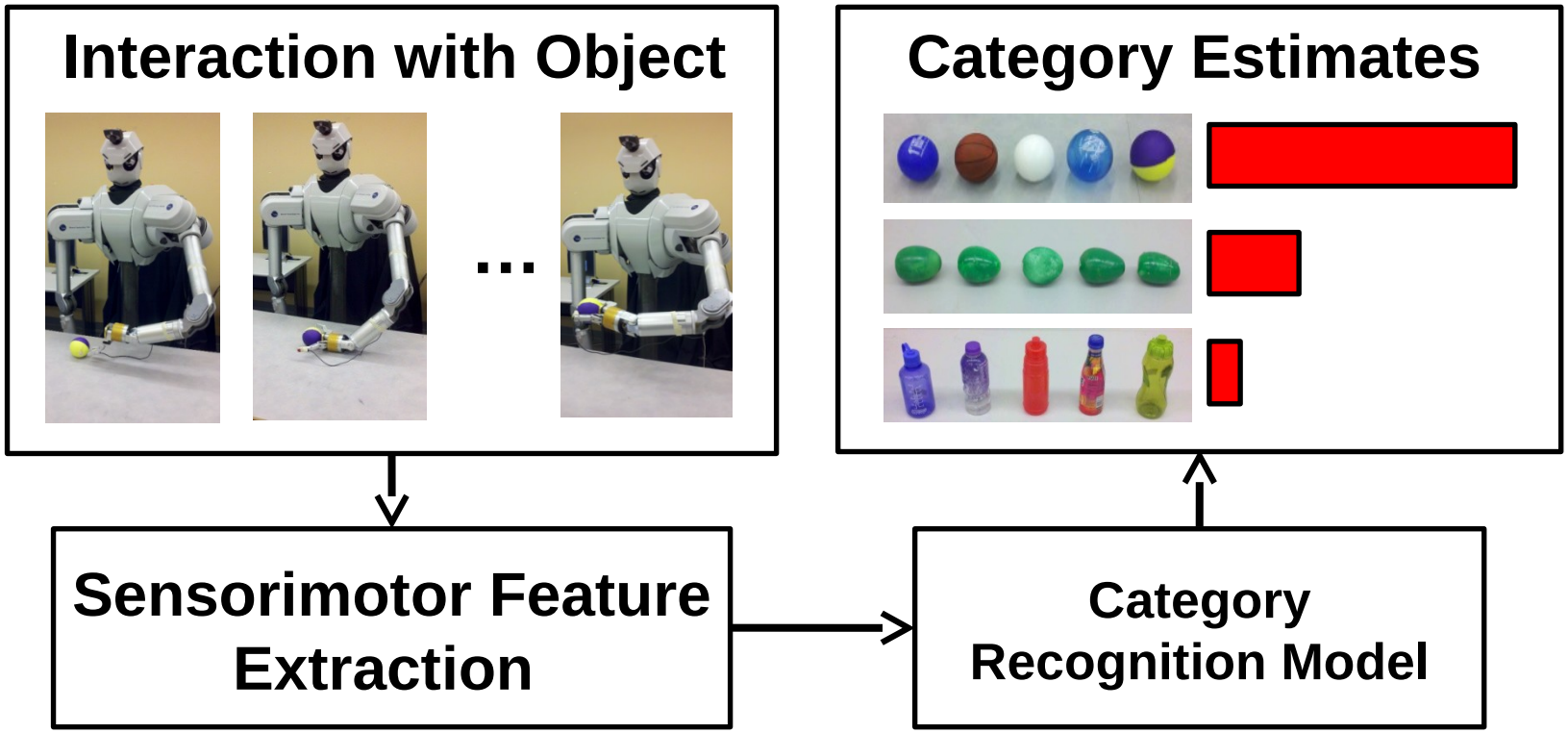
Coupling Action and Perception



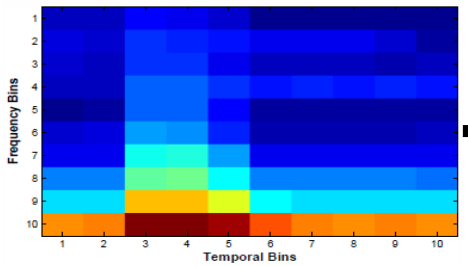
Sensorimotor Contexts

	audio (DFT)	haptics (joint torques)	proprioception (finger pos.)	Color	Optical flow	SURF
look				✓		✓
grasp	✓	✓	✓		✓	✓
lift	✓	✓			✓	✓
hold	✓	✓			✓	✓
shake	✓	✓			✓	✓
drop	✓	✓			✓	✓
tap	✓	✓			✓	✓
poke	✓	✓			✓	✓
push	✓	✓			✓	✓
press	✓	✓			✓	✓

Overview


Interaction with Object
...
Category Estimates
Sensorimotor Feature Extraction
Category Recognition Model

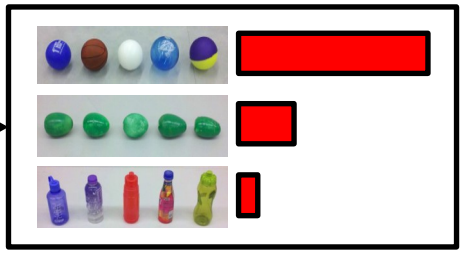
Context-specific Category Recognition



Observation from *poke-audio* context

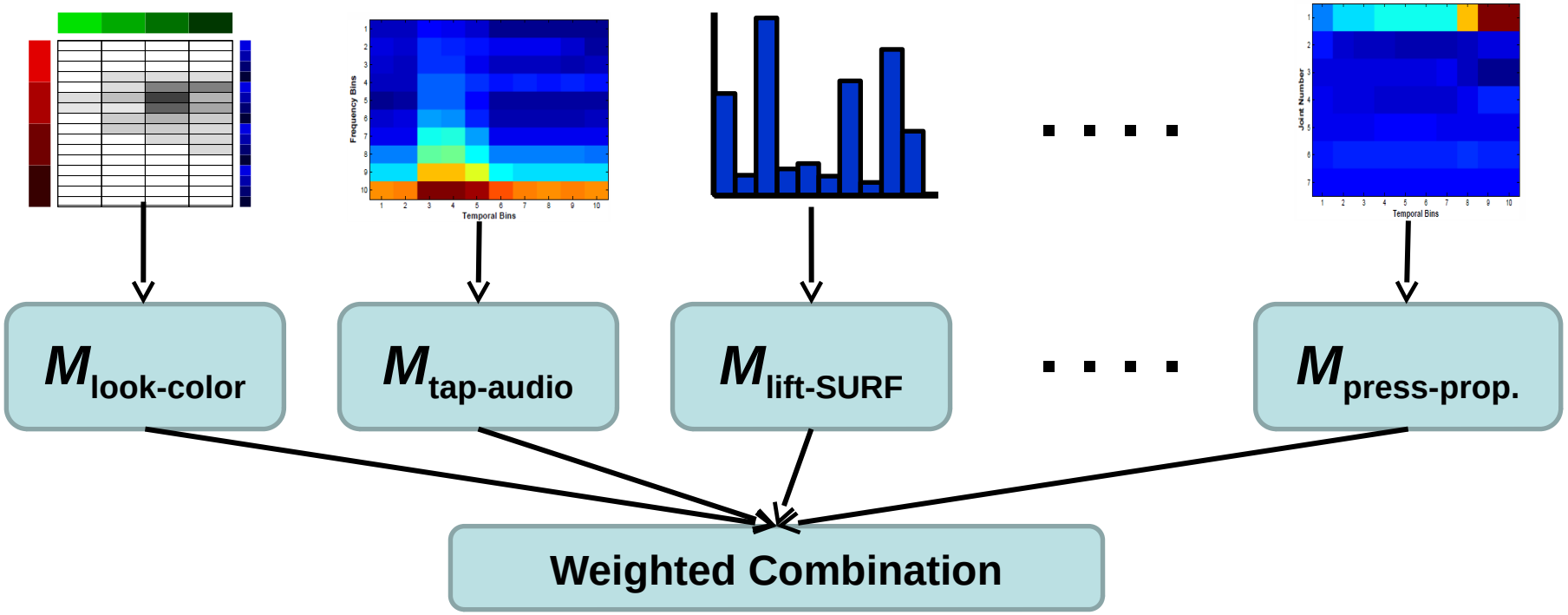
$M_{\text{poke-audio}}$

Recognition model for *poke-audio* context



Distribution over category labels

Combining Model Outputs



Deep Models for Non-Visual Perception

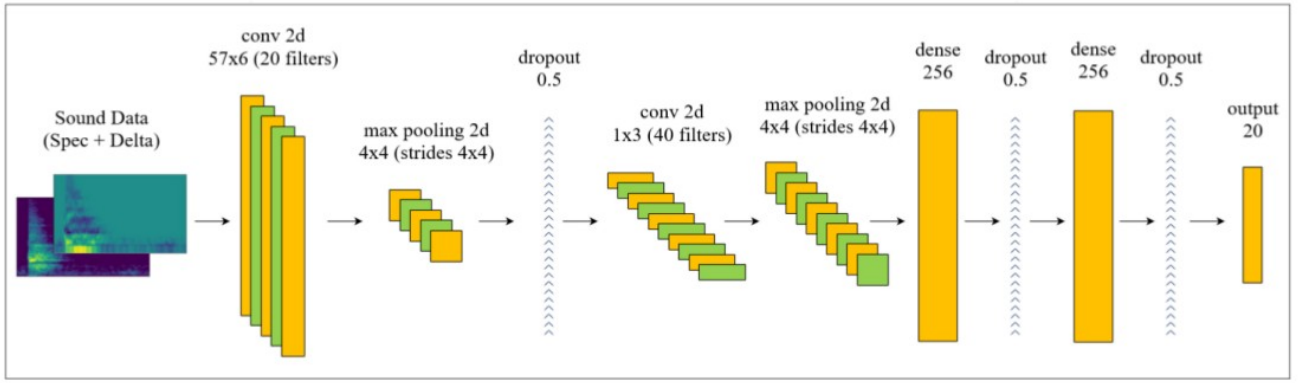


Fig. 2. The architecture of CNN used for sound classification.

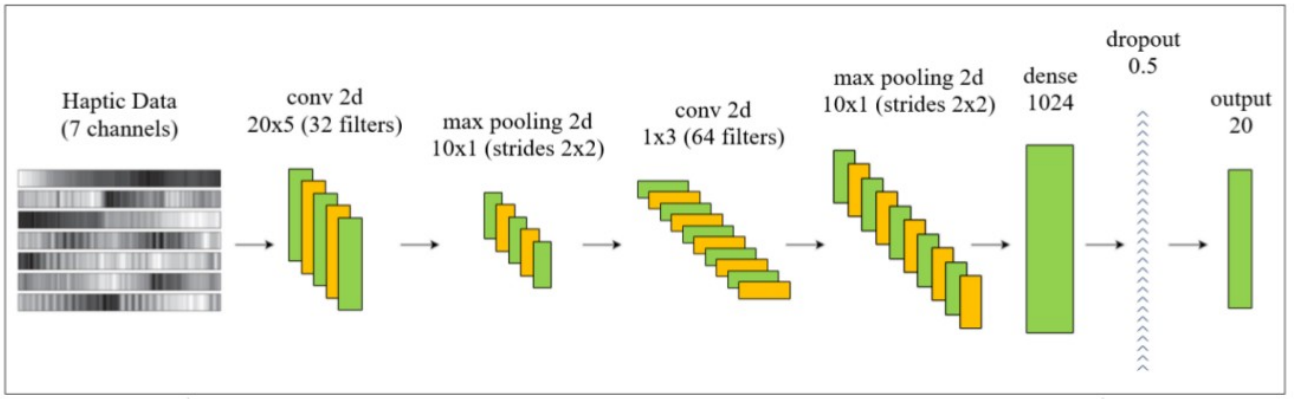


Fig. 3. The architecture of CNN used for haptic classification.

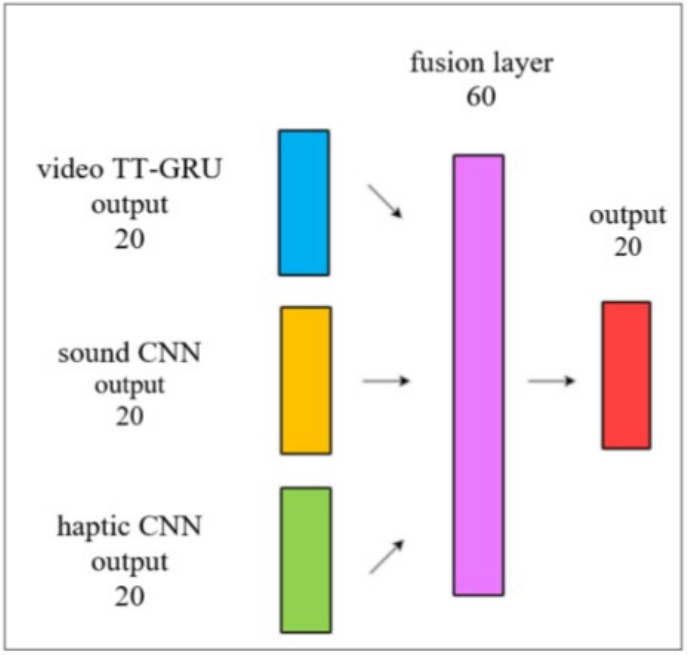
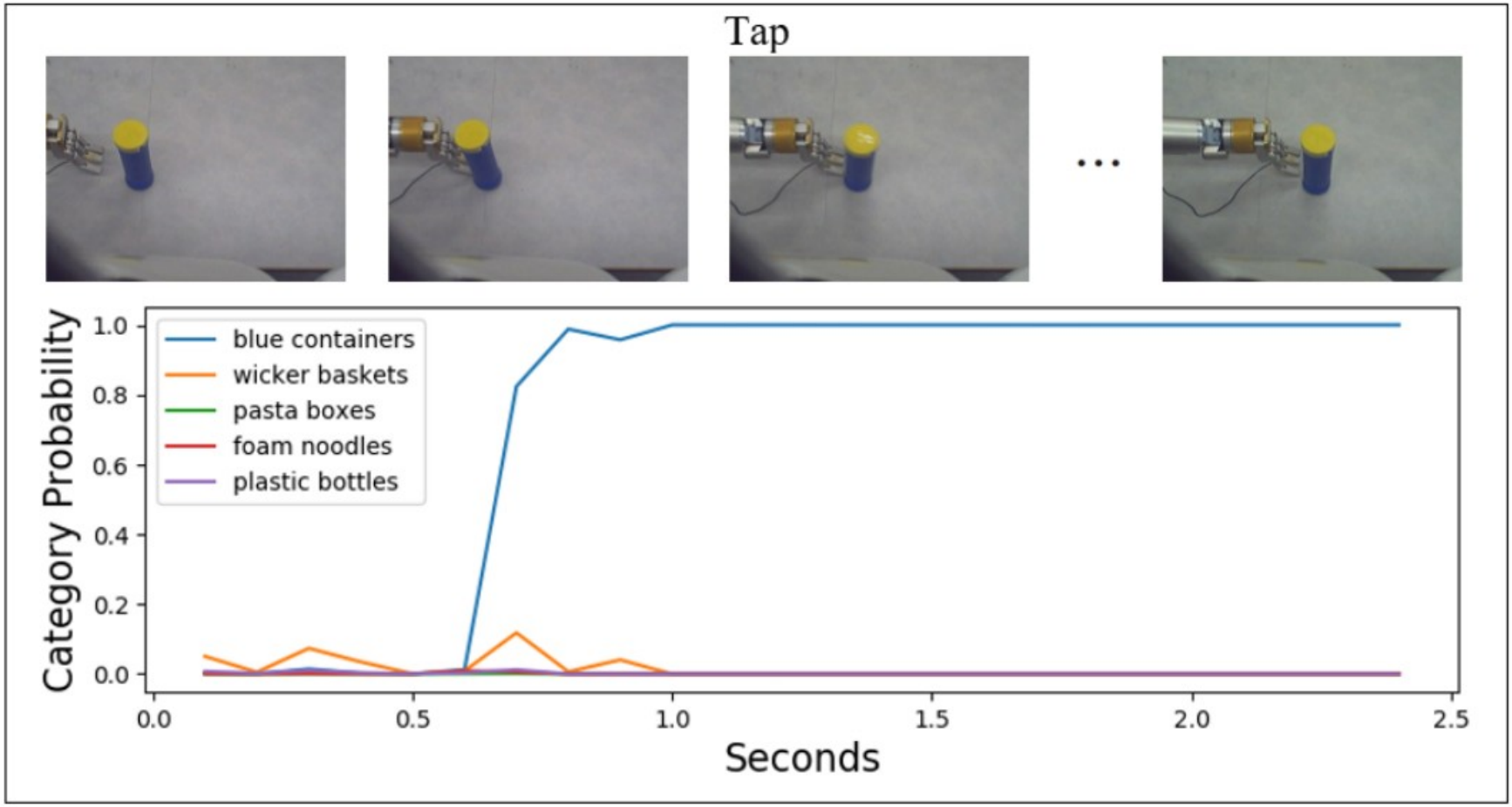


Fig. 4. The architecture multimodal network.

Tatiya, G., and Sinapov, J. (2019)
Deep Multi-Sensory Object Category Recognition Using Interactive Behavioral Exploration
2019 IEEE International Conference on Robotics and Automation (ICRA), Montreal, Canada, May 20-24, 2019.

Deep Models for Non-Visual Perception



Tatiya, G., and Sinapov, J. (2019)
Deep Multi-Sensory Object Category Recognition Using Interactive Behavioral Exploration
2019 IEEE International Conference on Robotics and Automation (ICRA), Montreal, Canada, May 20-24, 2019.

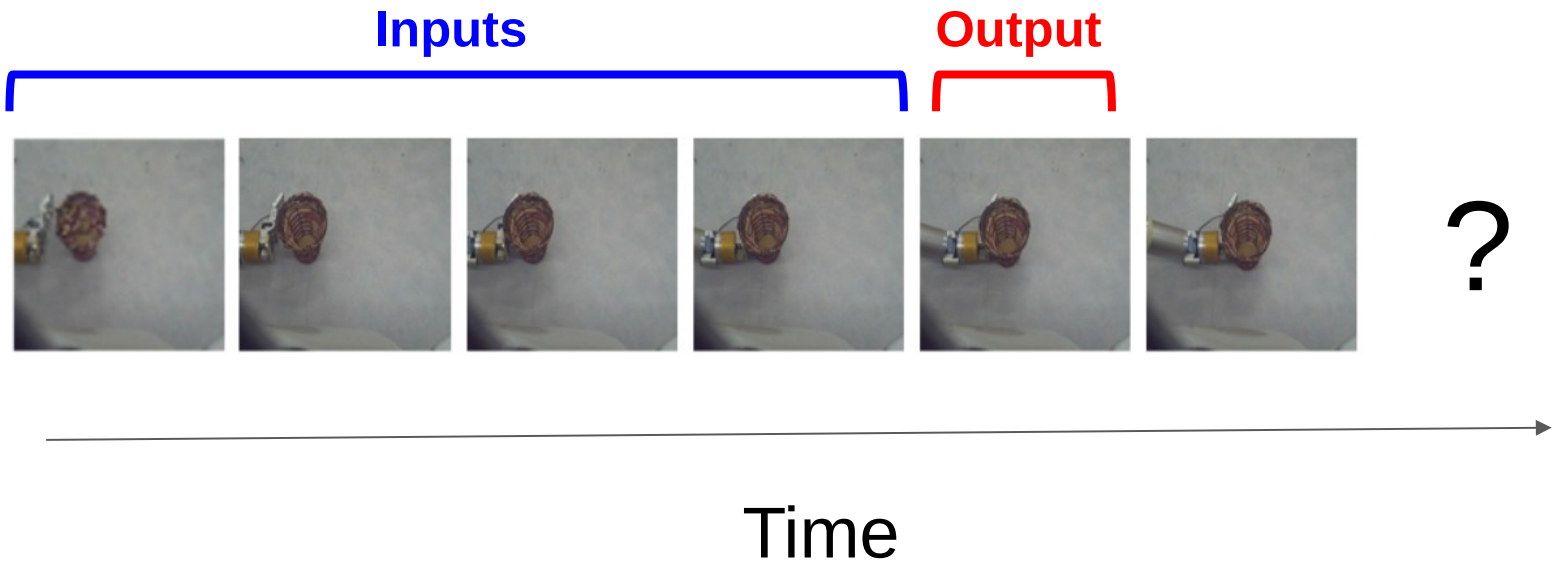
Visual Foresight

?

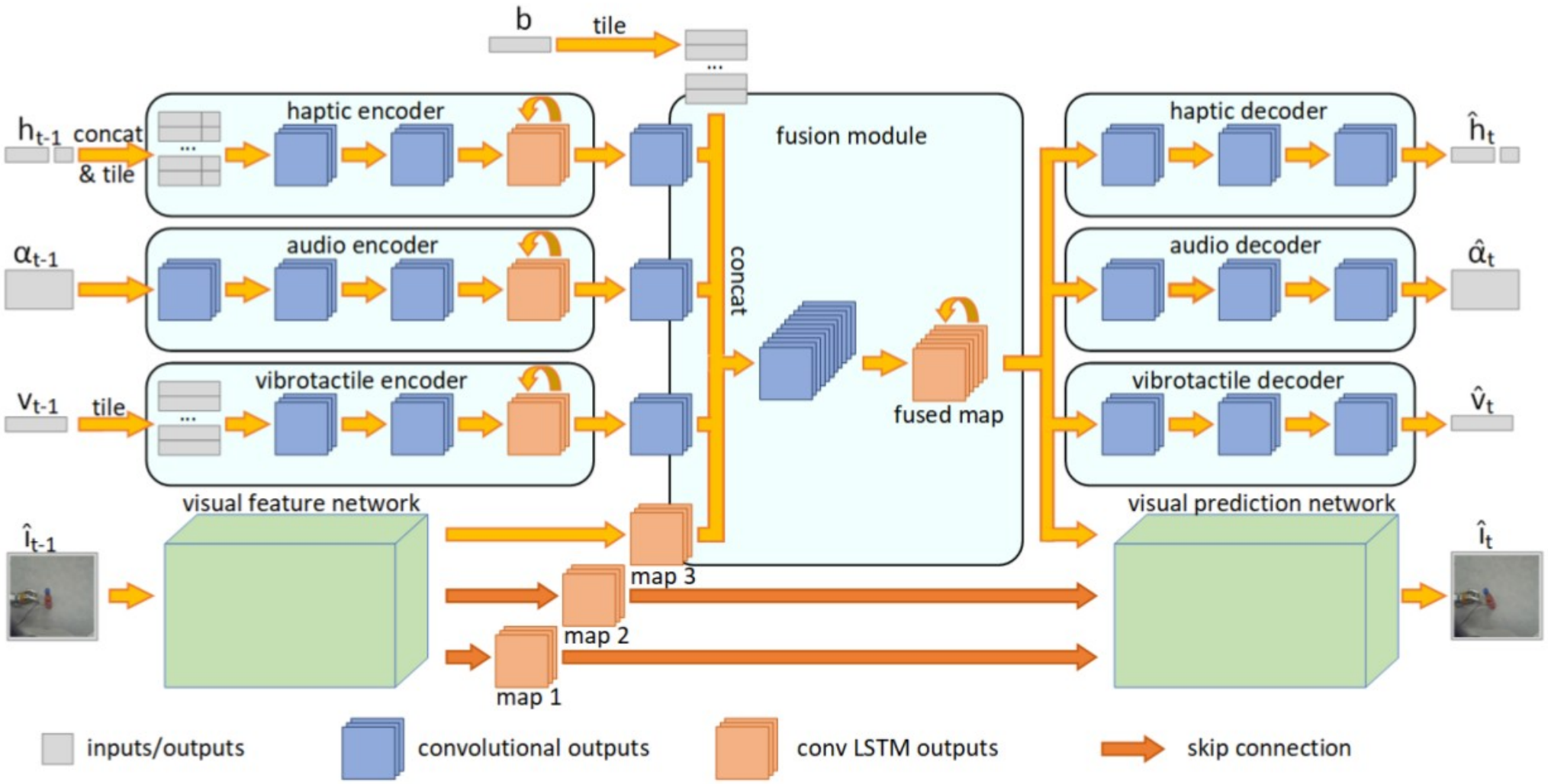
Time

Hosseini, R., Chen, X., Panetta, K., and Sinapov, J. (2021)
A Framework for Multisensory Foresight for Embodied Agents
In proceedings of the IEEE International Conference on Robotics and Automation (ICRA)

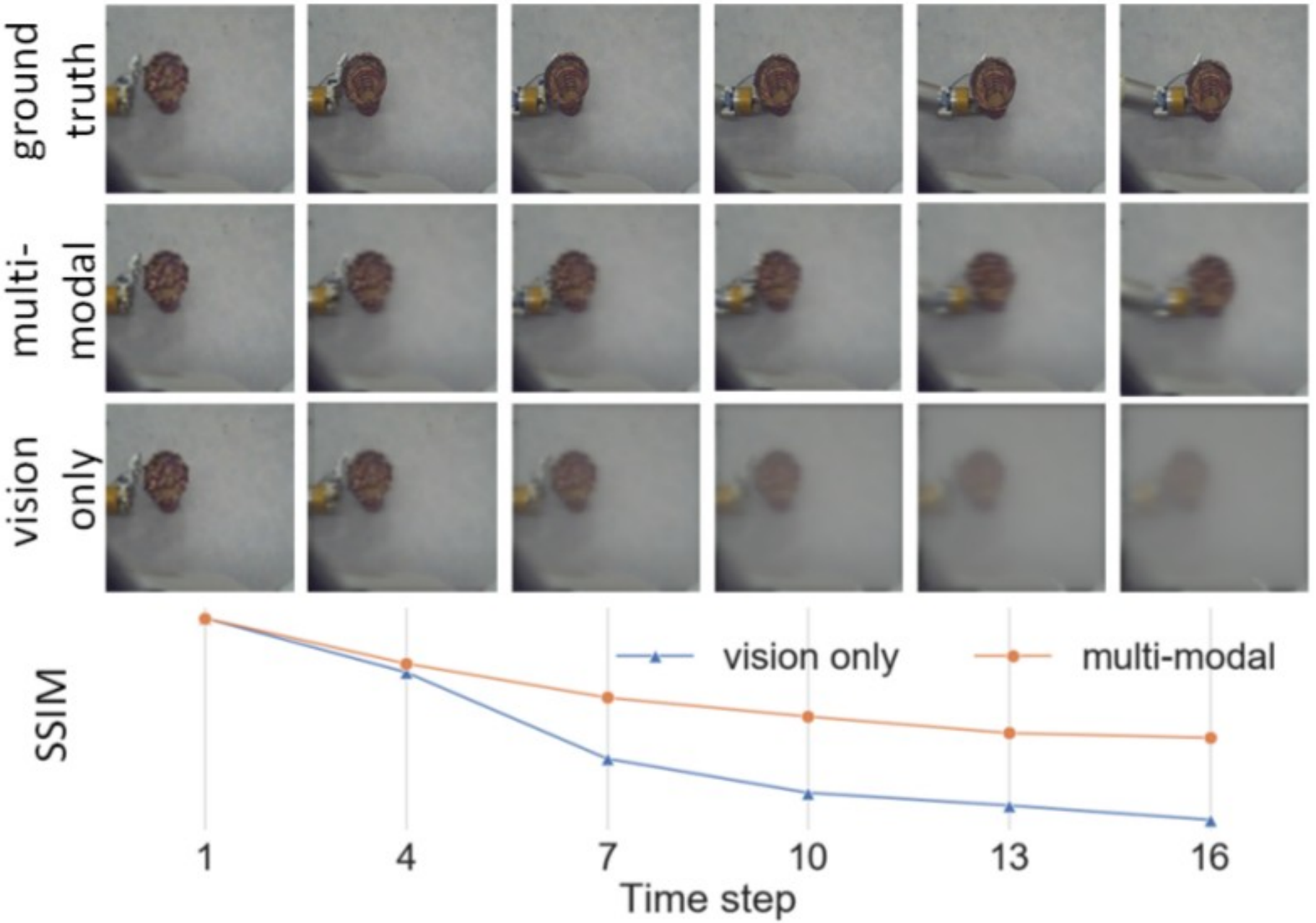
Visual Foresight



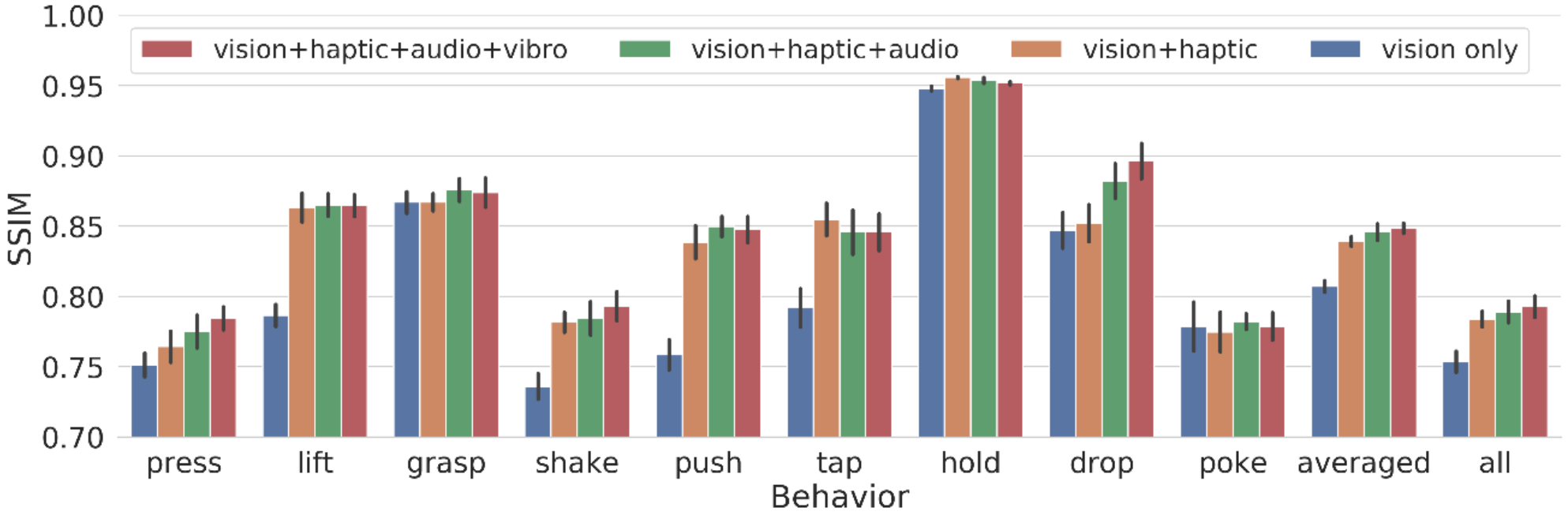
Hosseini, R., Chen, X., Panetta, K., and Sinapov, J. (2021)
A Framework for Multisensory Foresight for Embodied Agents
In proceedings of the IEEE International Conference on Robotics and Automation (ICRA)



Hosseini, R., Chen, X., Panetta, K., and Sinapov, J. (2021)
A Framework for Multisensory Foresight for Embodied Agents
In proceedings of the IEEE International Conference on Robotics and Automation (ICRA)



Hosseini, R., Chen, X., Panetta, K., and Sinapov, J. (2021)
A Framework for Multisensory Foresight for Embodied Agents
In proceedings of the IEEE International Conference on Robotics and Automation (ICRA)



Hosseini, R., Chen, X., Panetta, K., and Sinapov, J. (2021)
A Framework for Multisensory Foresight for Embodied Agents
In proceedings of the IEEE International Conference on Robotics and Automation (ICRA)

Take-home Message

- Behaviors allow robots not only to affect the world, but also to perceive it
- Non-visual sensory feedback improves object classification and perception tasks that are typically solved using vision alone
- A diverse sensorimotor repertoire is necessary for scaling up object recognition, categorization, and individuation to a large number of objects

Student-led Paper Presentation