Symbol Emergence in Robotics:
Language Acquisition via Real-world Sensorimotor Information

Tadahiro Taniguchi
1) Professor, College of Information Science & Engineering,
Ritsumeikan University
2) Visiting General Chief Scientist, AI solution center,
Panasonic

Invited talk at Gatsby-Kakenhi Joint Workshop on AI and Neuroscience
12th May 2017

Tadahiro Taniguchi @tanichu

- Professor, Emergent System Laboratory,
  College of Information Science and Engineering,
  Ritsumeikan University, Japan
  - 2003-2006: PhD student, Kyoto University
  - 2005-2008: JSPS research fellow, Kyoto University
  - 2008: Assistant professor, Ritsumeikan University
  - 2010: Associate professor, Ritsumeikan University
  - 2015-2016: Visiting Associate Professor,
    Imperial College London
  - **2017: Professor, Ritsumeikan University**
  - **2017: Visiting General Chief Scientist,**
    Panasonic Corporation
    AI solution center (20% C.A.)
- **Research Topics**
  - Machine learning, Intelligent Robotics,
    Symbol emergence in robotics, Language acquisition

Contents

Computational Understanding of Mental Development

From Behavioral Learning to Language Acquisition

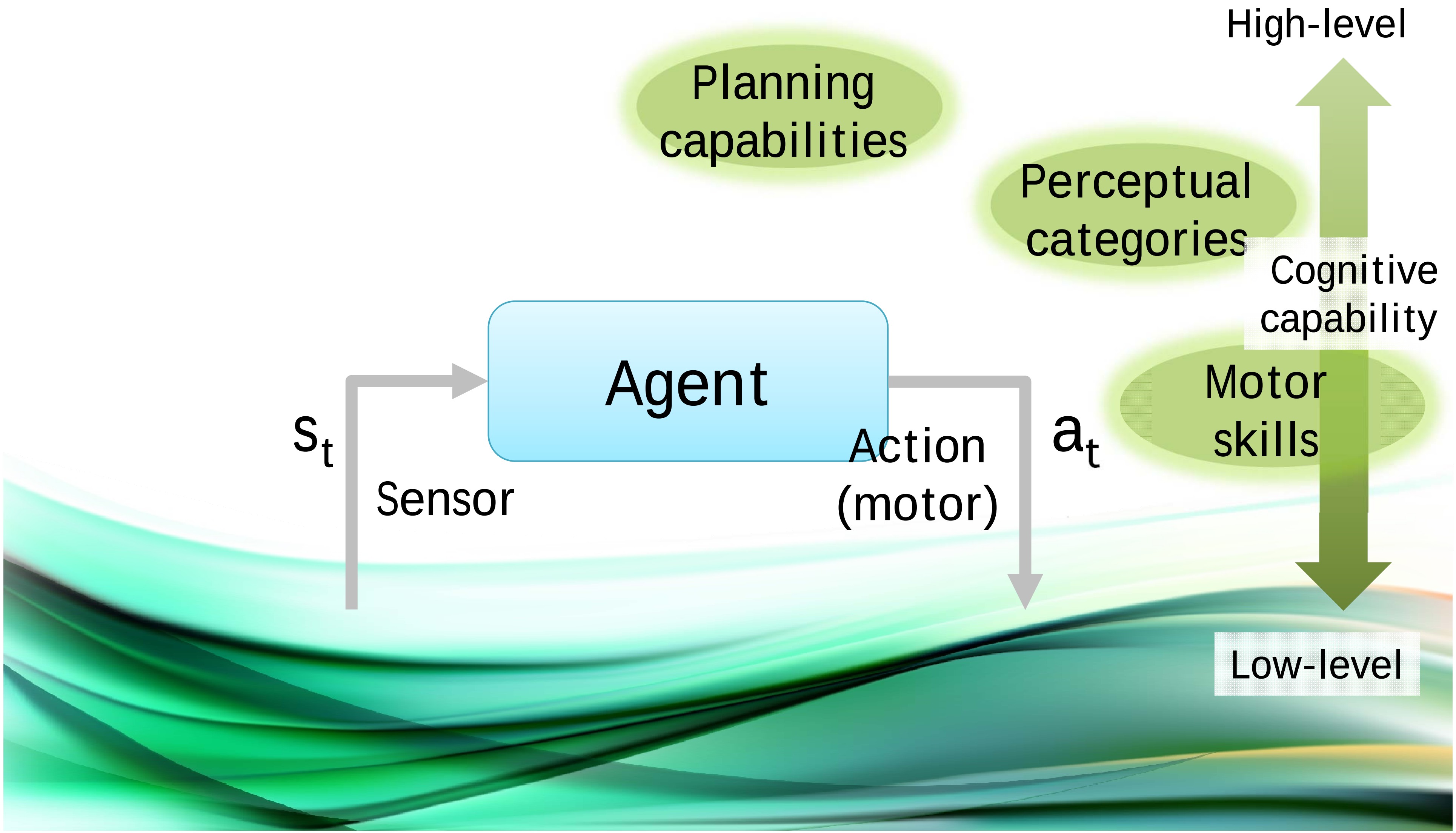
- A human child acquires many physical skills, concepts, and knowledge, including language, through physical and social interaction with his/her environment.
- How do we become able to communicate via symbols?
- We'd like to obtain an understanding of the **computational process** of mental development and language acquisition.

Constructive approach

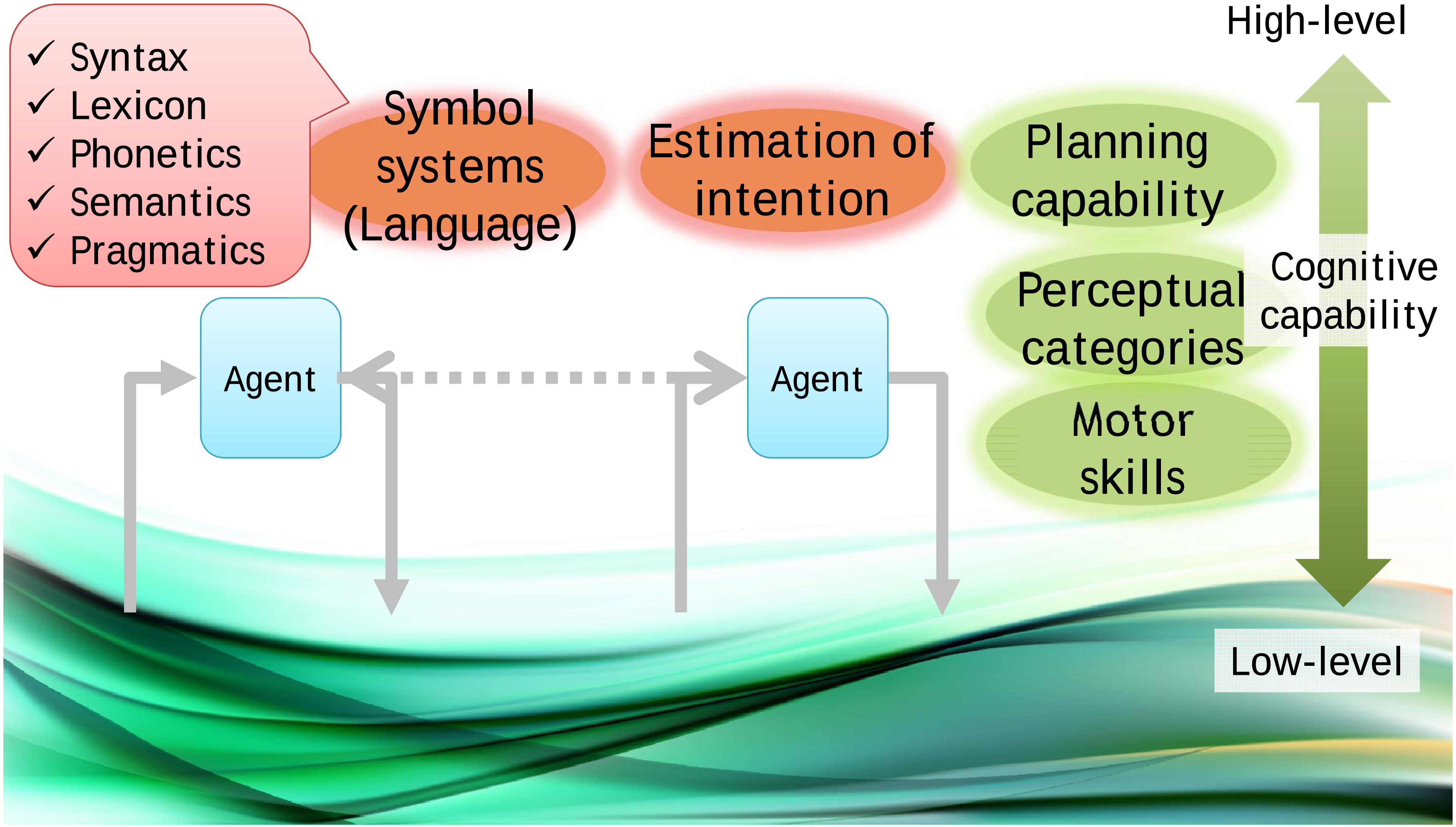
Develop robotic and computational models to better understand the original

<u>Symbol Emergence in Robotics</u>

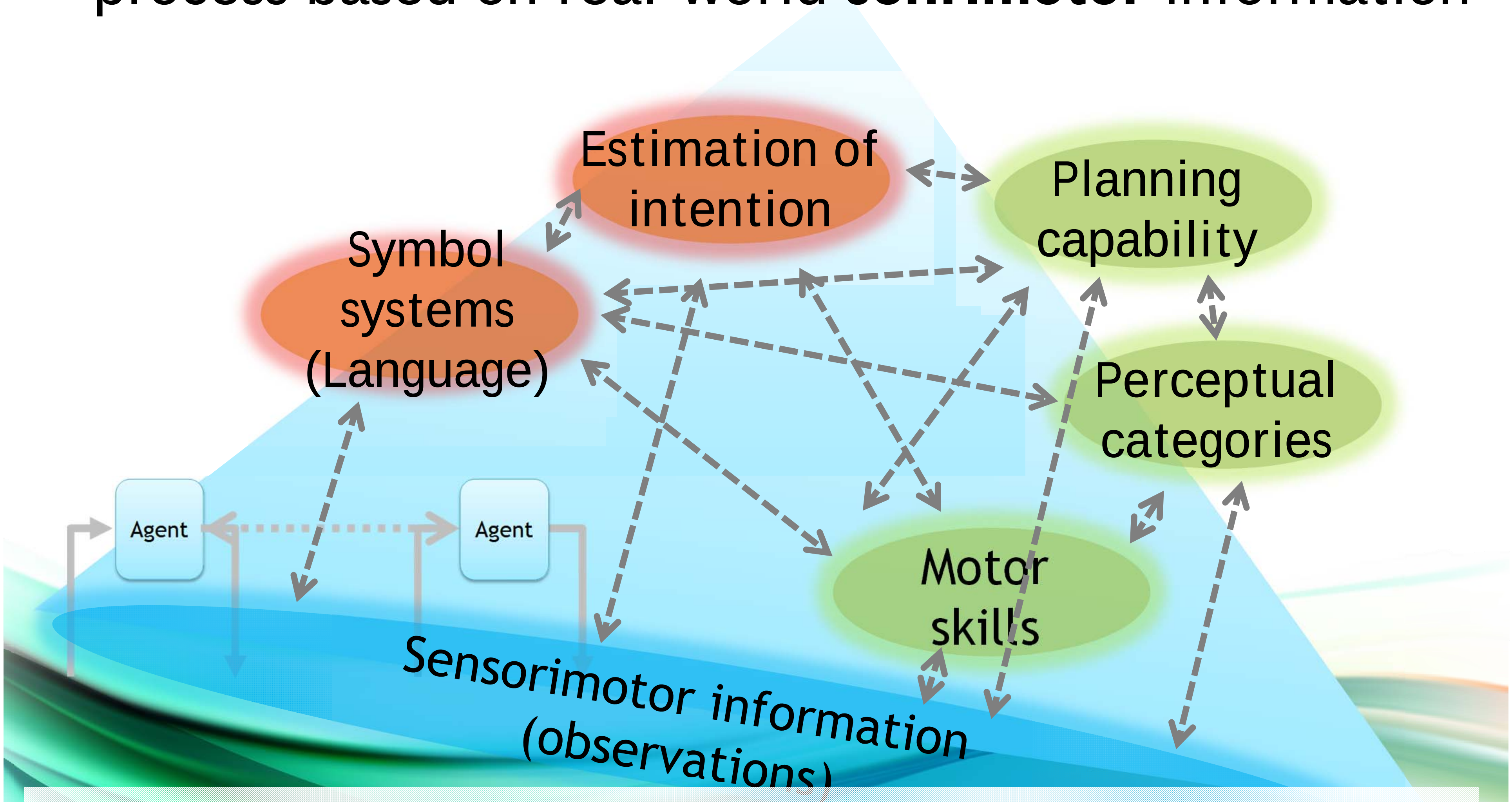
Development though a self-organizational learning process based on real-world **senrimotor** information



Development though a self-organizational learning process based on real-world **senrimotor** information



Development though a self-organizational learning process based on real-world **senrimotor** information



✓ **Bottom-up organization** of a variety of mutually dependent cognitive functions based on sensorimotor information

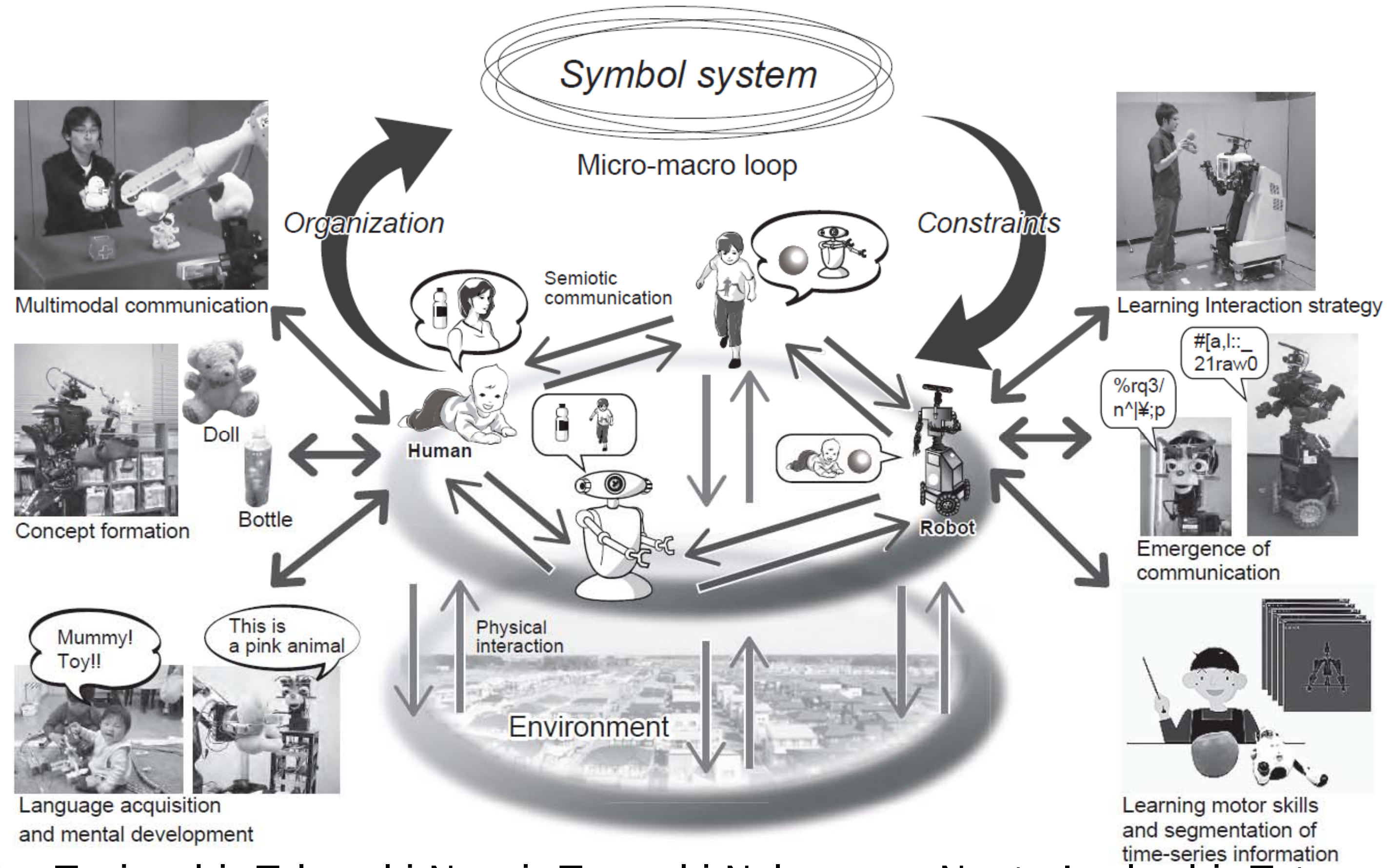
ADVANCED ROBOTICS, 2016
http://dx.doi.org/10.1080/01691864.2016.1164622

RSJ Taylor & Francis
Taylor & Francis Group

SURVEY PAPER

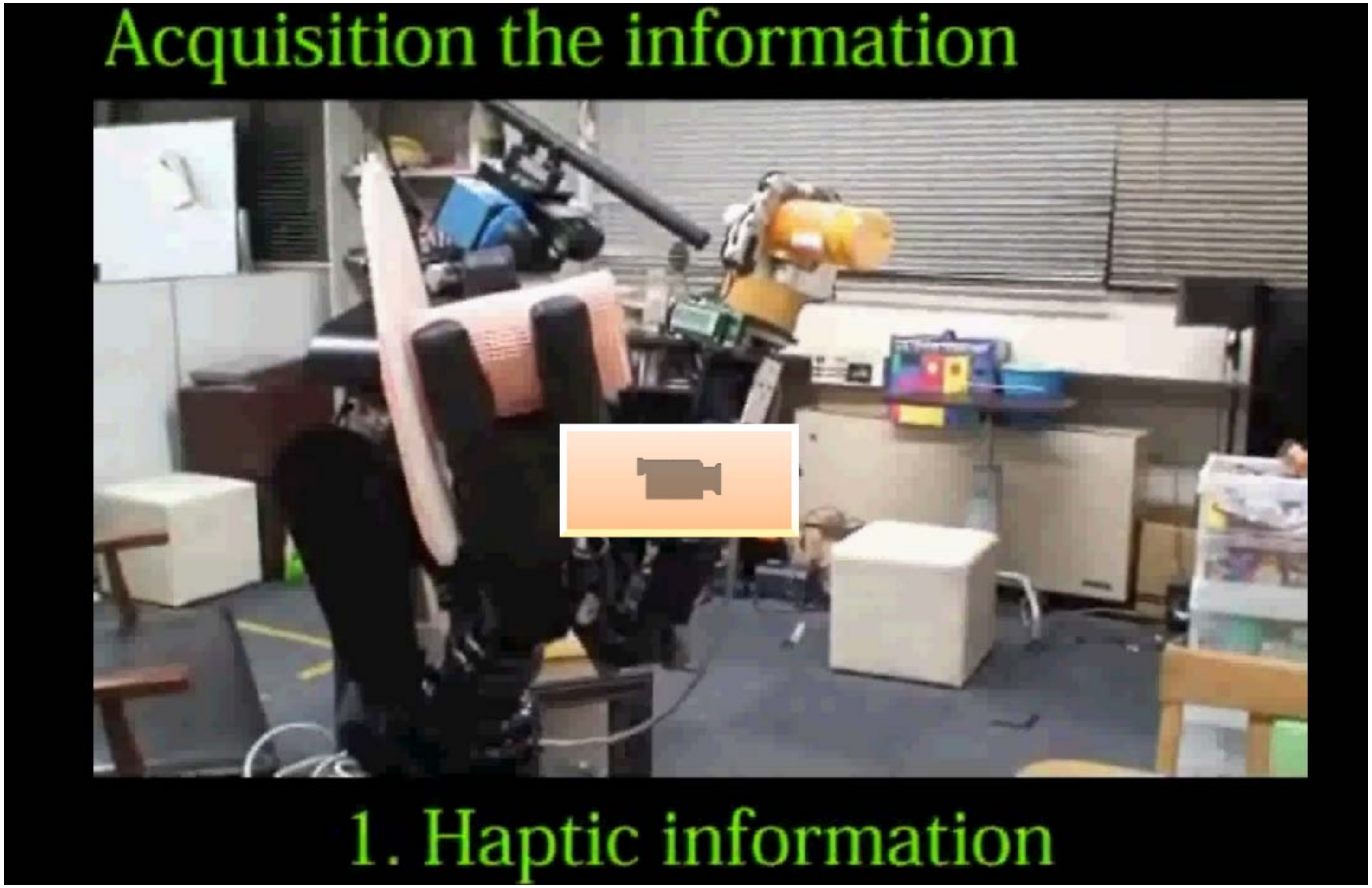
Symbol emergence in robotics: a survey

Tadahiro Taniguchi[a], Takayuki Nagai[b], Tomoaki Nakamura[b], Naoto Iwahashi[c], Tetsuya Ogata[d] and Hideki Asoh[e]



Tadahiro Taniguchi, Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata, and Hideki Asoh, Symbol Emergence in Robotics: A Survey
Advanced Robotics, .(2016)DOI:10.1080/01691864.2016.1164622

Multimodal Categorization and Lexical Acquisition by an Autonomous Robot [Nakamura+ 2009-]

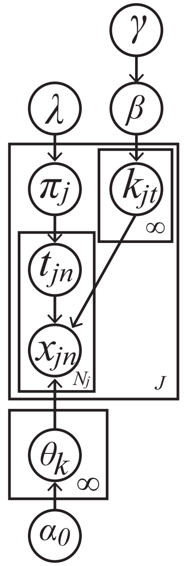


Takaya Araki, Tomoaki Nakamura, Takayuki Nagai, Shogo Nagasaka, **Tadahiro Taniguchi**, Naoto Iwahashi. Online Learning of Concepts and Words Using Multimodal LDA and Hierarchical Pitman-Yor Language Model. IEEE/RSJ International Conference on Intelligent Robots and Systems 2012 (IROS 2012), 1623-1630 .(2012)

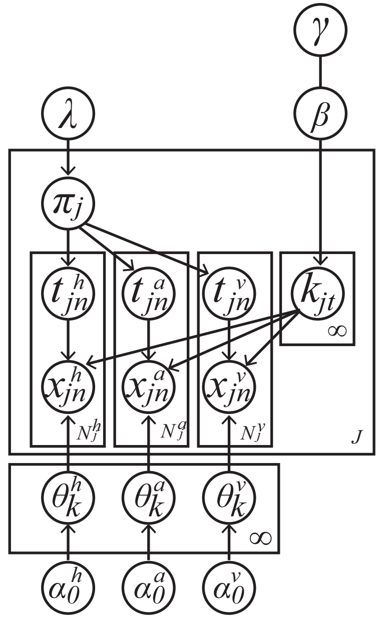
Multimodal latent Dirichlet allocation(MLDA) / Hierarchical Dirichlet Processes(MHDP)

[Nakamura+ 2009, 2011]

- The **MLDA** is a multimodal categorization method that is an extension of the LDA [Blei+ 2014].
- The MLDA was originally proposed for making a **robot** form **object categories** in an unsupervised manner.
- **Multimodal Hierarchical Dirichlet process (MHDP)** is a nonparametric Bayesian extension of MLDA [Nakamura+ 2011].



HDP
[Teh+ 2006]



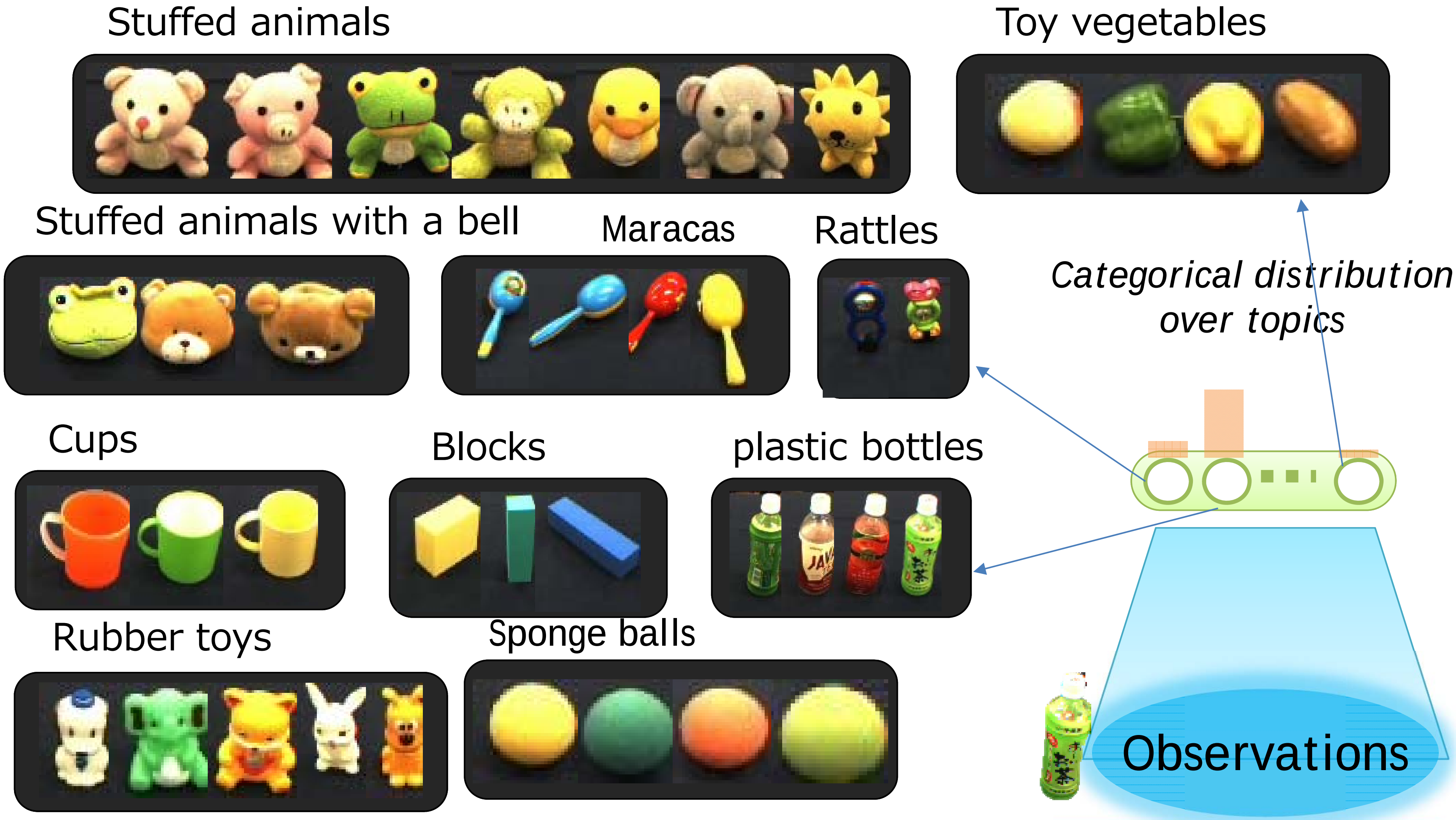
MHDP
[Nakamura+ 2011]

	Observations	Latent variable
LDA/HDP	**Words** in a **document** (i.e., Bag of words)	**Topic**
MLDA/ MHDP	Multimodal (**visual**, **auditory**, and **haptic**) **features** obtained from an **object** (i.e., Bag of features)	**Object category**

[Teh+ 2006] Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei. Hierarchical dirichlet processes. Journal of the American Statistical Association, 101(476):1566-1581, 2006.

[Nakamura+ 2011] Tomoaki Nakamura, Takayuki Nagai, and Naoto Iwahashi. Multimodal categorization by hierarchical Dirichlet process. In IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 1520-1525, 2011.

Categorization result based on real-world multimodal sensorimotor information

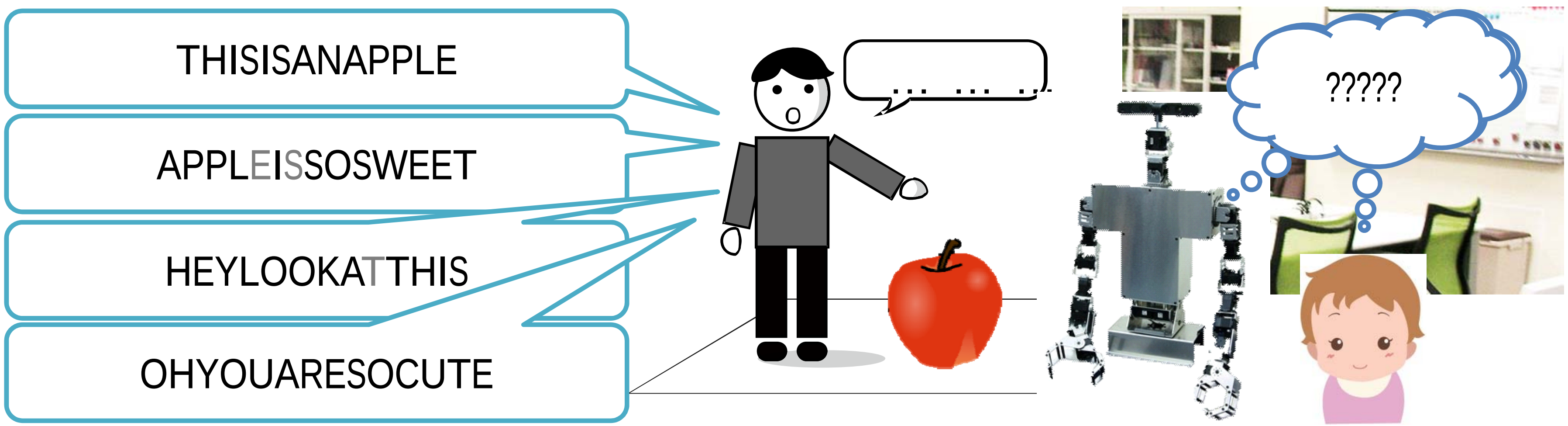


By integrating multimodal information, the robot formed categories represented by latent variables that were similar to most of the human participants.

Contents

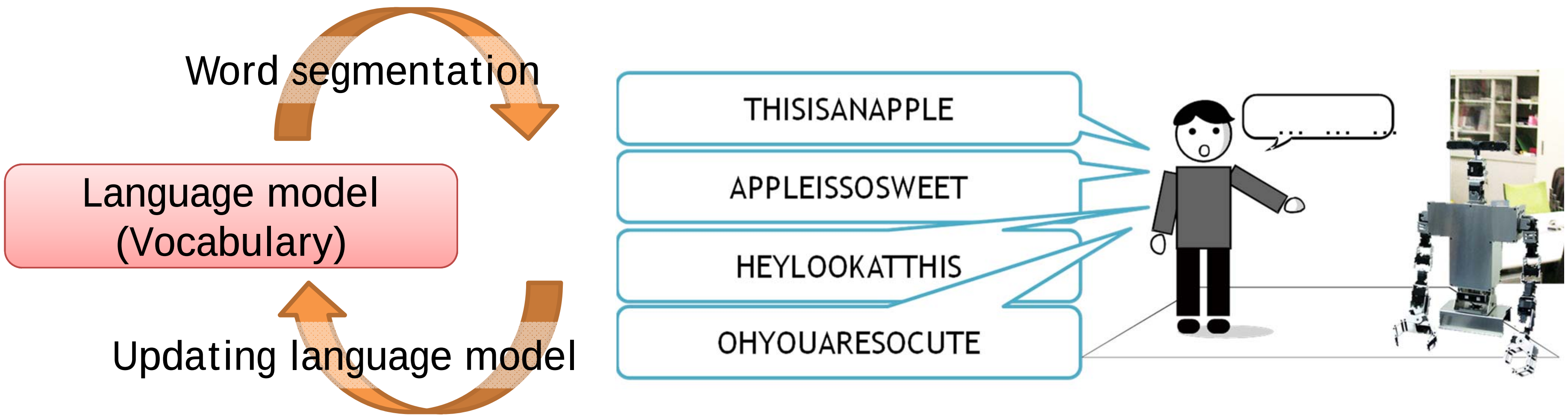
Word discovery (segmentation) in language acquisition

- When parents speak to their children, they rarely use "isolated words," but use continuous word sequences, i.e. sentences.
- Word and phoneme discovery (segmentation) is a primary task of language acquisition.
- The child has to perform word segmentation without pre-existing knowledge of vocabulary because children do not know lists of words before they learn.



Unsupervised word segmentation

- Unsupervised word segmentation
  - No preexisting dictionaries are used.
  - Input data is test data.
  - A nonparametric Bayesian framework for word segme [Goldwater+ 09]
  - Unsupervised word segmentation method based on the Nested Pitman–Yor language model (NPYLM) [Mochihashi+ 09].



S. Goldwater, T. L. Griffiths, and M. Johnson, "A Bayesian framework for word segmentation: exploring the effects of context.," *Cognition*, vol. 112, no. 1, pp. 21–54, 2009.

Daichi Mochihashi, Takeshi Yamada, Naonori Ueda."Bayesian Unsupervised Word Segmentation with Nested Pitman-Yor Language Modeling", ACL-IJCNLP 2009, pp.100-108, 2009.

Analysis

"Alice in Wonderland"の解析

first,shedreamedoflittlealiceherself,andonceagainthetinyhandswereclaspedup
onherknee,andthebrighteagereyeswerelookingupintohersshecouldhearthevery
tonesofhervoice,andseethatqueerlittletossofherheadtokeepbackthewanderingh
airthatwouldalwaysgetintohereyesandstillasshelistened,orseemedtolisten,thew
holeplacearoundherbecamealivethestrangecreaturesofherlittlesister'sdream.the
longgrassrustledatherfeetasthewhiterabbithurriedbythefrightenedmousesplashe
dhiswaythroughtheneighbouringpoolshecouldheartherattleoftheteacupsasthema
rchhareandhisfriendssharedtheirneverendingmeal,andtheshrillvoiceofthequeen…

first, she dream ed of little alice herself ,and once again the tiny hand s were
clasped upon her knee ,and the bright eager eyes were looking up into hers --
shecould hearthe very tone s of her voice , and see that queer little toss of
herhead to keep back the wandering hair that would always get into hereyes --
and still as she listened , or seemed to listen , thewhole place a round her
became alive the strange creatures of her little sister 'sdream. thelong grass
rustled ather feet as thewhitera bbit hurried by -- the frightened mouse splashed
his way through the neighbour ing pool -- shecould hearthe rattle ofthe tea cups
as the marchhare and his friends shared their never -endingme a l ,and the …

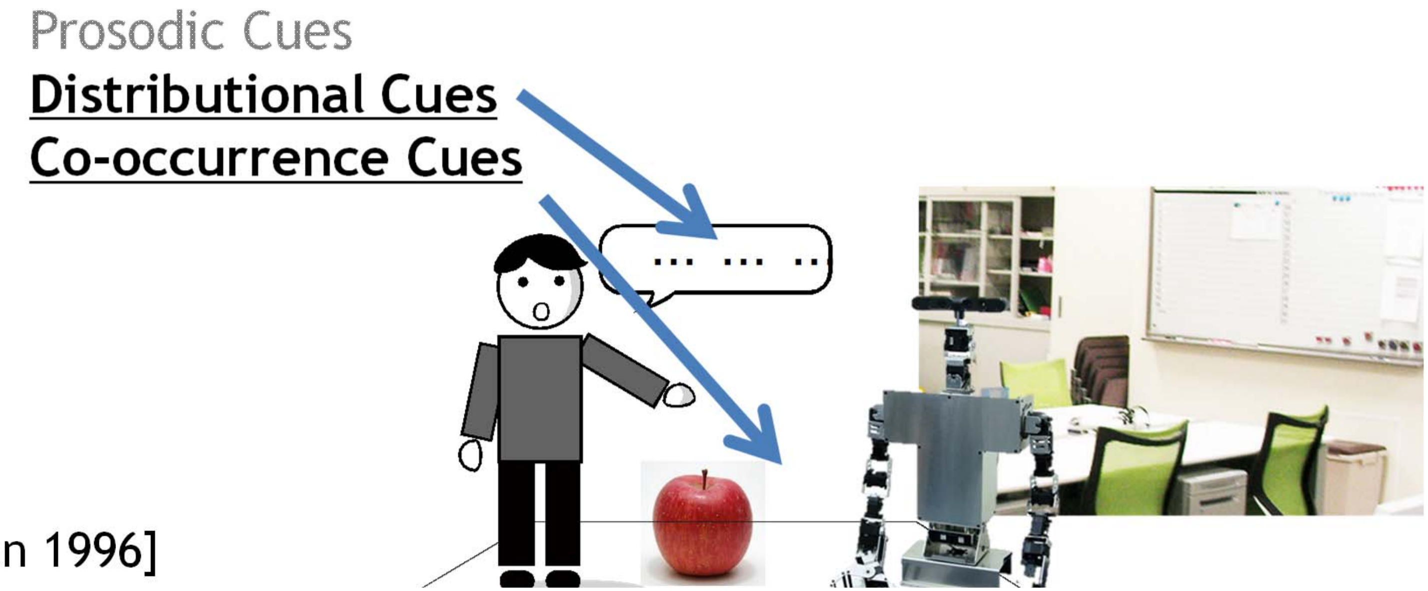
[From Mochihashi's presentation slide]

http://chasen.org/~daiti-m/paper/jfssa2009segment.pdf

Challenges in
Real-world Unsupervised Word Discovery Tasks

- NPYLM presumes that the target document (sentences) is transcribed without **errors**. If there are **phoneme recognition errors**, its performance becomes dramatically worse.

A) Mitigating negative effects of phoneme recognition errors
B) Learning a phoneme system (acoustic model).
C) Grounding discovered words



[Saffuran 1996]

Contents

1. Introduction
2. Lexical acquisition tasks
   A) Direct phoneme and word discovery from speech signals
   B) Simultaneous acquisition of word units and multimodal categories
   C) Online spatial concept acquisition
3. Future challenges

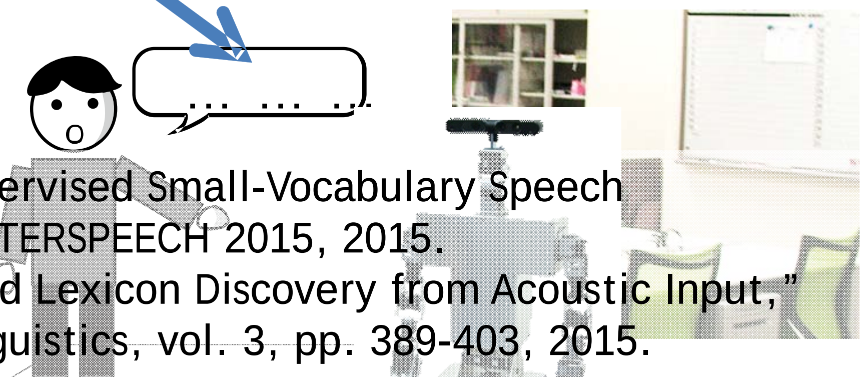
Simultaneous acquisition of **phoneme** and **language models**

- ✓ In most related studies about unsupervised word discovery, they used a pre-existing phoneme model and did not make a robot learn a phoneme system (an acoustic model).
- ✓ There are still few studies about unsupervised simultaneous learning of phoneme and language models from speech signals [Kamper+ 15, Lee+ 15].
- ✓ It has been suggested that the learning processes of acoustic and language models (phoneme and lexical systems) are mutually dependent.

Prosodic Cues

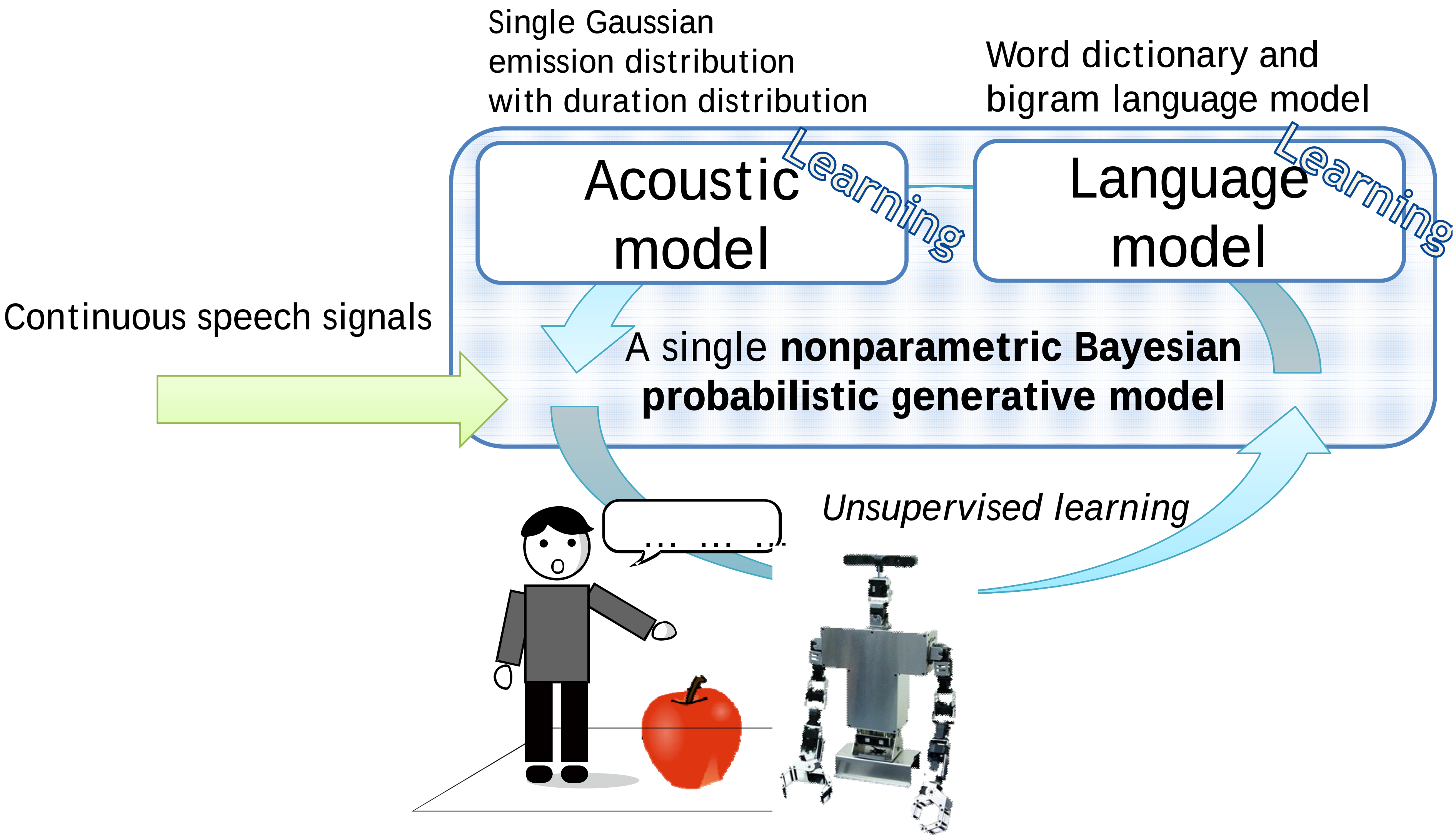
Making full use of the **Distributional Cues** directly from speech signals

Co-occurrence Cues



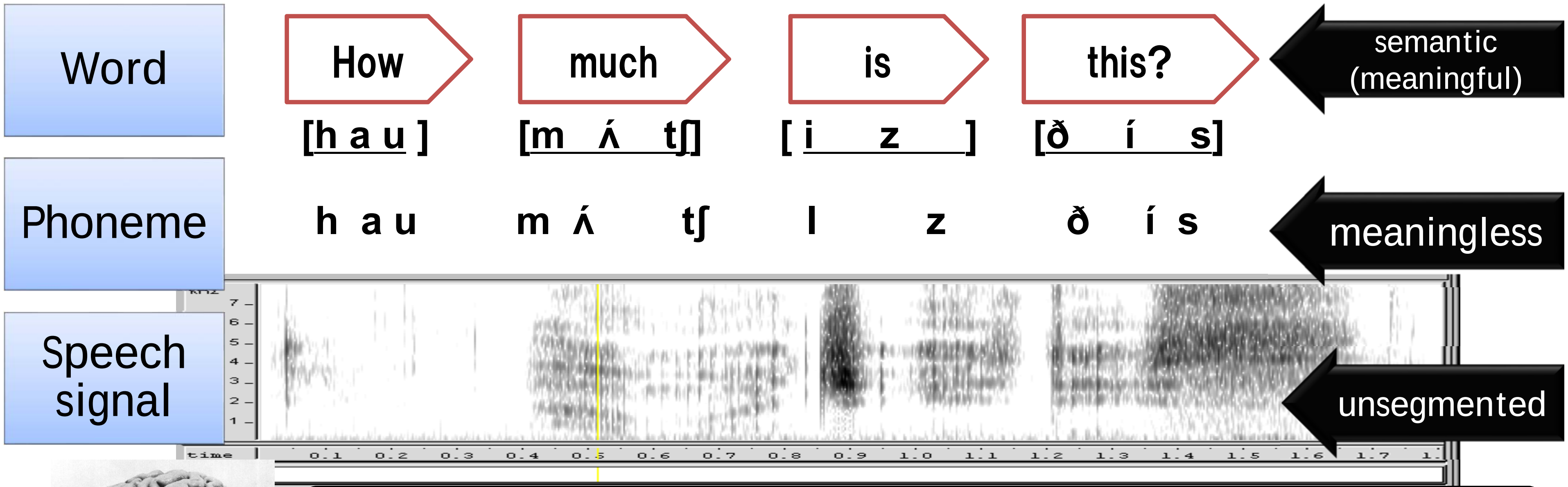
H. Kamper, A. Jansen, and S. Goldwater, "Fully Unsupervised Small-Vocabulary Speech Recognition Using a Segmental Bayesian Model," in INTERSPEECH 2015, 2015.
C.-y. Lee, T. J. O. Donnell, and J. Glass, "Unsupervised Lexicon Discovery from Acoustic Input," Transactions of the Association for Computational Linguistics, vol. 3, pp. 389-403, 2015.

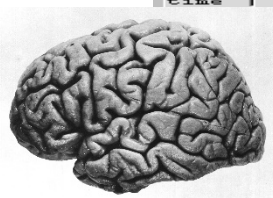
Direct phoneme and word discovery from speech signals



Double articulation structure in semiotic data

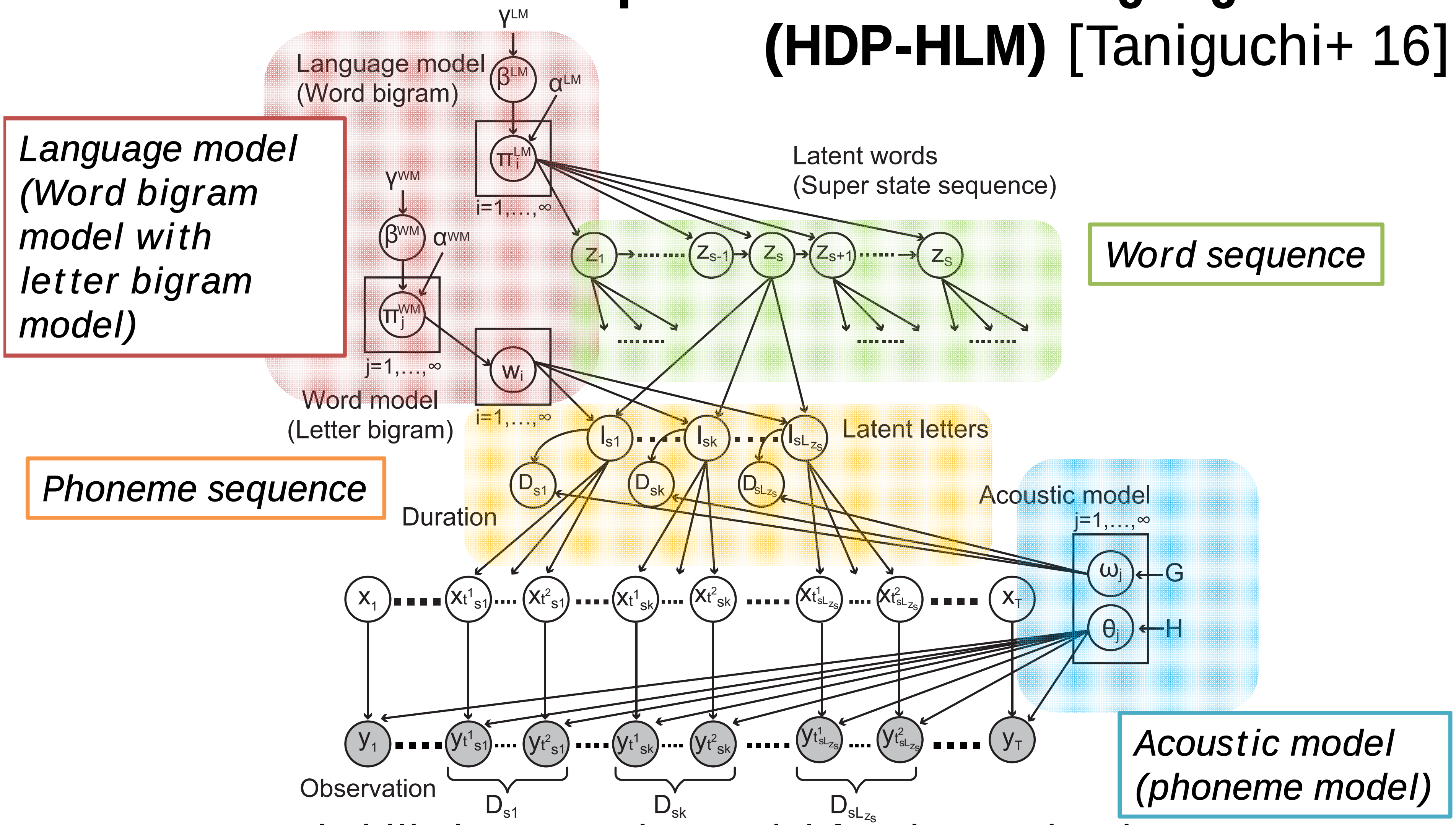
- Semiotic time-series data often has double articulation
  - Speech signal is a continuous and high-dimensional time-series.
  - Spoken sentence is considered a sequence of **phonemes.**
  - Phonemes are grouped into **words**, and people give them meanings.





Does the human brain have a special capability to analyze double articulation structures embedded in time-series data?

Hierarchical Dirichlet process hidden language model (HDP-HLM) [Taniguchi+ 16]

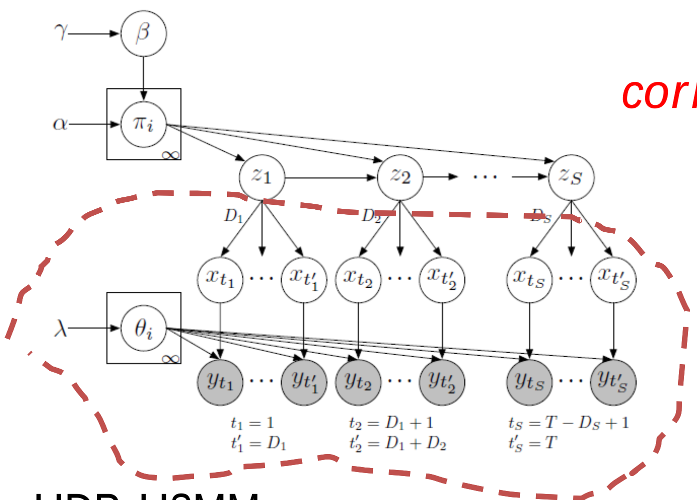


<u>A probabilistic generative model for time-series data having double articulation structure</u>

<u>**Tadahiro Taniguchi**</u>, Shogo Nagasaka, Ryo Nakashima, Nonparametric Bayesian Double Articulation Analyzer for Direct Language Acquisition from Continuous Speech Signals, IEEE Transactions on Cognitive and Developmental Systems.(2016)

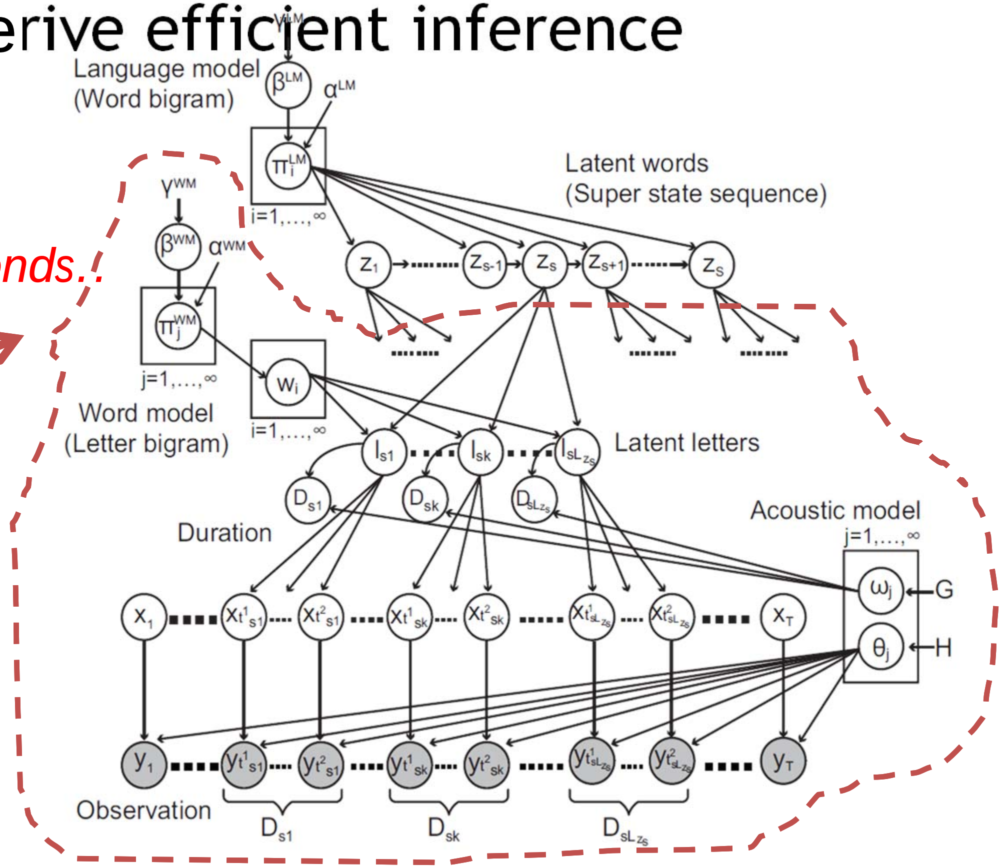
HDP-HLM as an extension of HDP-HSMM

- ✓ **HDP-HLM** can be regarded as an extension of **HDP-HSMM** [Johnson'13]
- ✓ This property helps us to derive efficient inference procedure.



HDP-HSMM
(hierarchical Dirichlet process hidden semi-Markov model)

corresponds..!



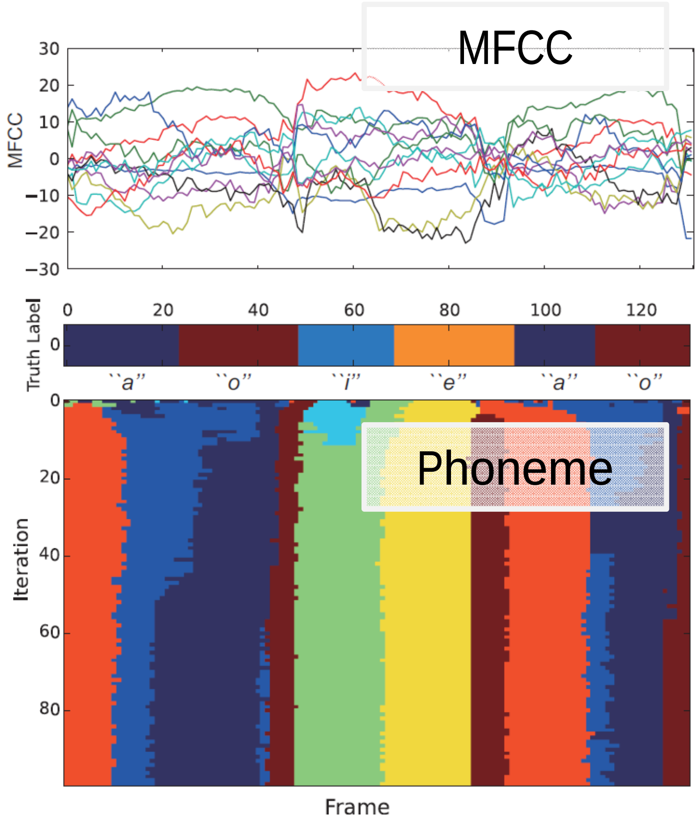
Matthew J Johnson and Alan S Willsky. Bayesian nonparametric hidden semi-markov models. *The Journal of Machine Learning Research*, Vol. 14, No. 1, pp. 673–701, 2013.

Evaluation experiment using artificial 2 or 3 words sentences with Japanese five vowels

- ✓ Five artificial words {aioi, aue, ao, ie, uo} prepared by connecting five Japanese vowels.
- ✓ 30 sentences (25 two-word and 5 three-word sentences) are prepared and each sentence is recorded twice by four Japanese speakers.
- ✓ MFCC (frame size =25ms, shift = 10ms, frame rate 100hz)

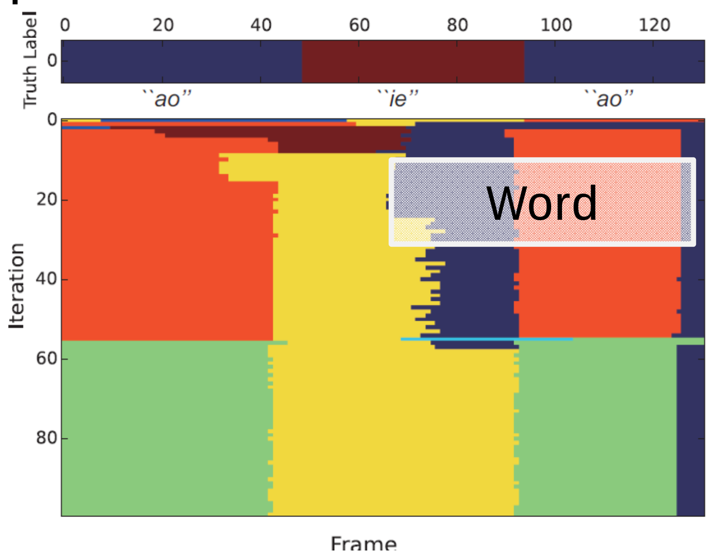
uo aue ie
ie ie uo
aue ao ie
ao ie ao
aioi uo ie

* HDP-HLM are trained separately for each speaker.



- ✓ The inference procedure could gradually estimate the boundaries of words and phonemes.

ex) ao-ie-ao



Iteration of Gibbs sampler

Experimental results

Table 1. ARI for estimated latent letters and words.

Method	Letter ARI	Word ARI	AM	LM
NPB-DAA with DSAE (MAP)	**0.589**	**0.705**		
NPB-DAA with DSAE	0.426	**0.398**		
NPB-DAA (MAP)	**0.612**	0.328		
NPB-DAA	0.551	0.359		
Conventional DAA	0.584	0.072		
Julius (GMM monophone + phoneme dictionary + NPYLM)	0.483	0.315	✓	
(GMM [illegible]	[illegible]	[illegible]6	✓	
(GMM [illegible]	[illegible]	[illegible]0	✓	
(GMM triphone + phoneme dictionary + latticelm)	0.269	0.203	✓	
Julius (DNN triphone + phoneme dictionary + NPYLM)	0.634	0.333	✓	
Julius (GMM monophone + word dictionary)	0.565	0.548	✓	✓
Julius (GMM triphone + word dictionary)	0.516	0.636	✓	✓
Julius (DNN triphone + word dictionary)	0.675	0.779	✓	✓

The NPB-DAA with DSAE even outperformed MFCC-based off-the-shelf speech recognition system.

Tadahiro Taniguchi, Ryo Nakashima, Hailong Liu and Shogo Nagasaka, Double Articulation Analyzer with Deep Sparse Autoencoder for Unsupervised Word Discovery from Speech Signals, Advanced Robotics.(2016)

Contents

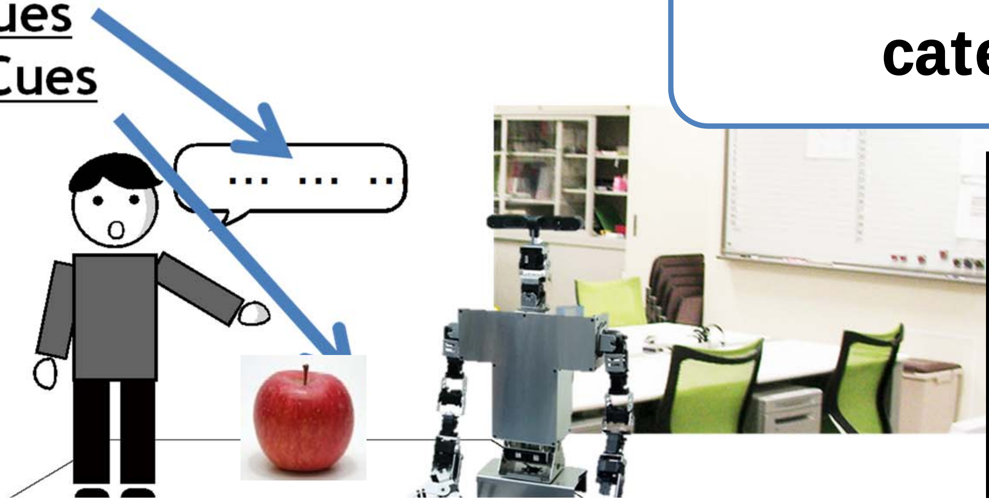
Making use of **co-occurrence cue**

- ✓ To detect the co-occurrence of an object and a phrase, the robot has to form the category of the object beforehand, or simultaneously.

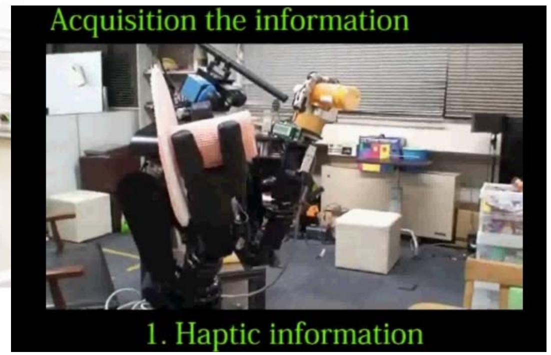
Example,  and, an *“apple”*

- ✓ **How** can a robot form “**object categories”** without knowing the names of objects?

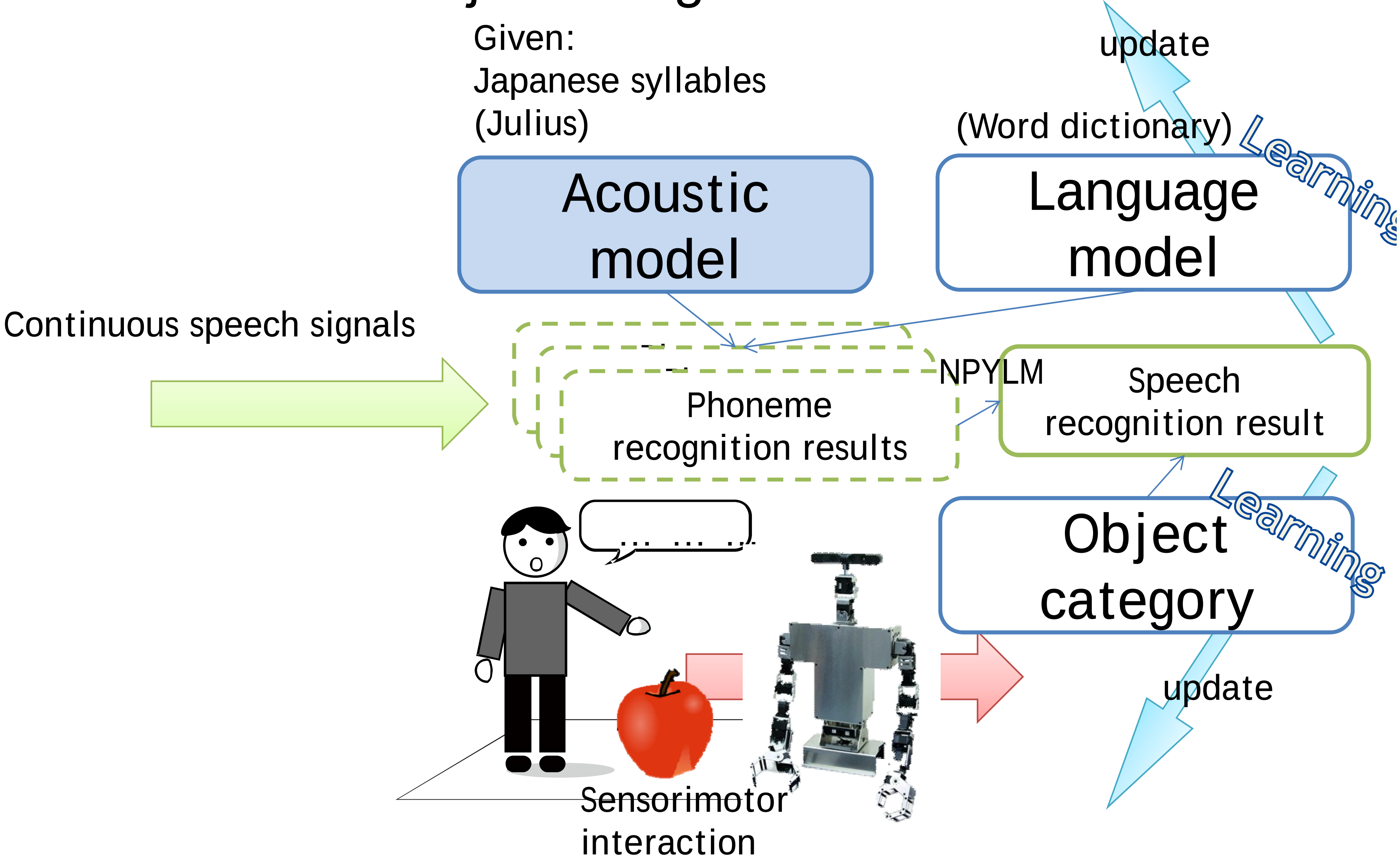
Prosodic Cues
<u>Distributional Cues</u>
<u>Co-occurrence Cues</u>



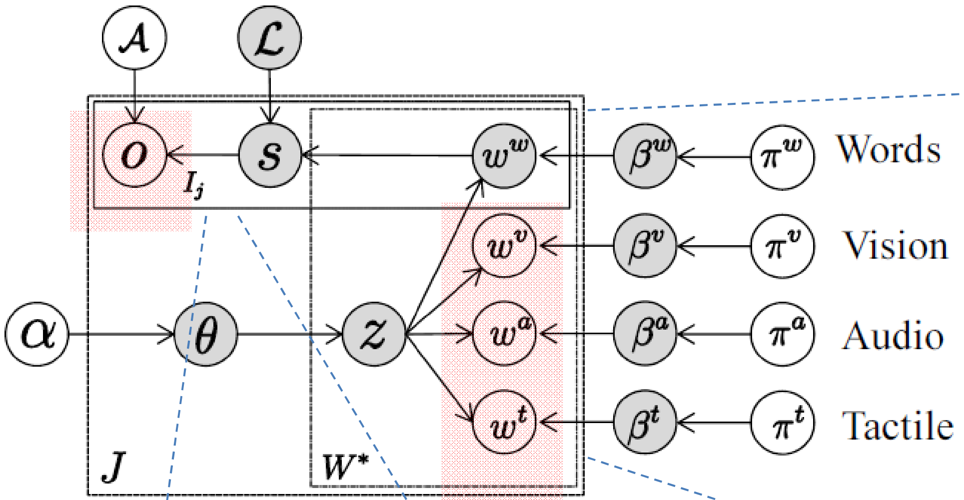
Multimodal object categorization



Simultaneous acquisition of word units and multimodal object categories

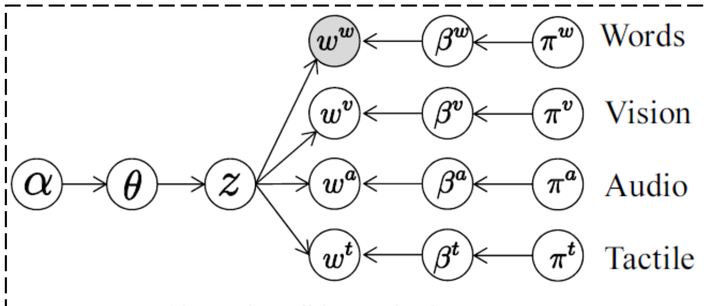


Probabilistic generative model for simultaneous acquisition of word units and multimodal object categories

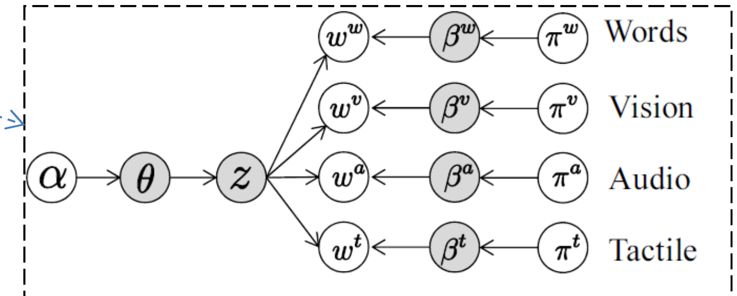


observations

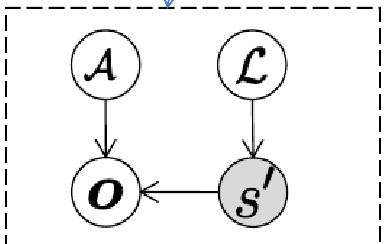
by MLDA

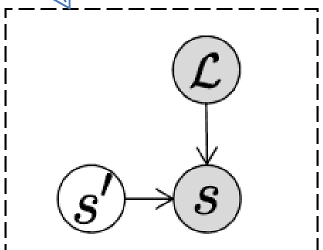


(c) Word candidates selection



(d) Object concept


(a) Speech recognition


(b) Language model

by Julius

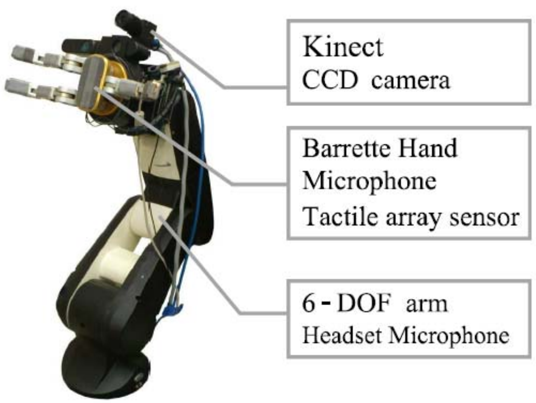
by NPYLM

Each component is updated using approximate MAP estimation iteratively.

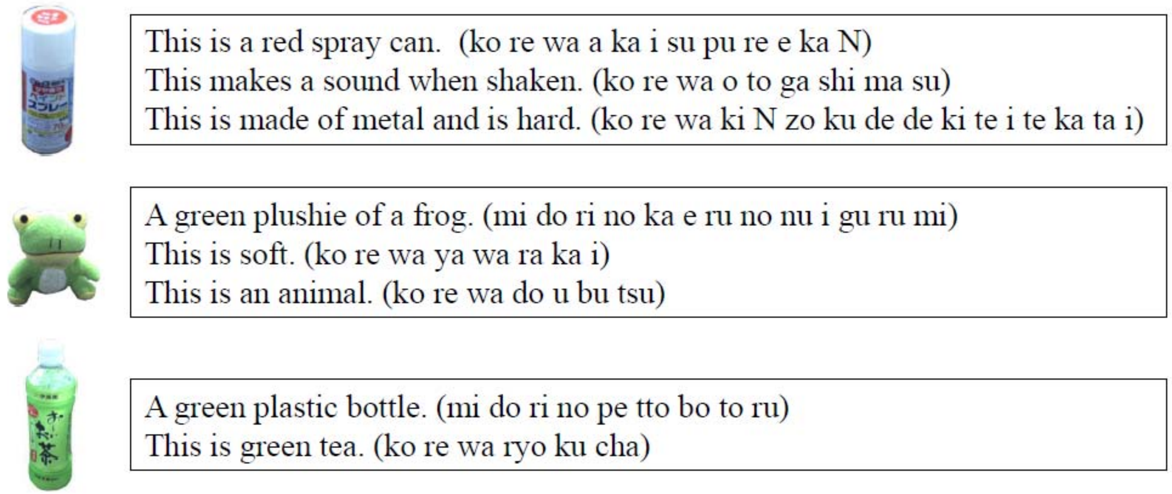
-> **Approximate inference procedure**

Tomoaki Nakamura, Takayuki Nagai, Kotaro Funakoshi, Shogo Nagasaka, **Tadahiro Taniguchi**, and Naoto Iwahashi, Mutual Learning of an Object Concept and Language Model Based on MLDA and NPYLM, 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'14), 600 - 607 .(2014)

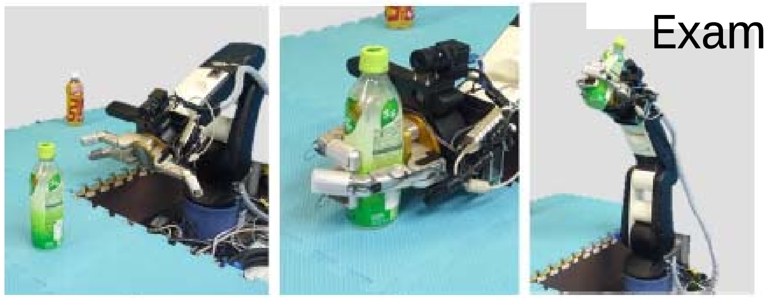
Overview of experiment and results



Robot

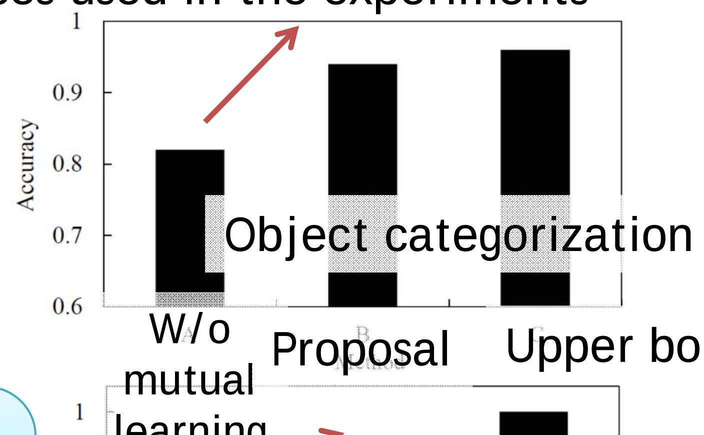


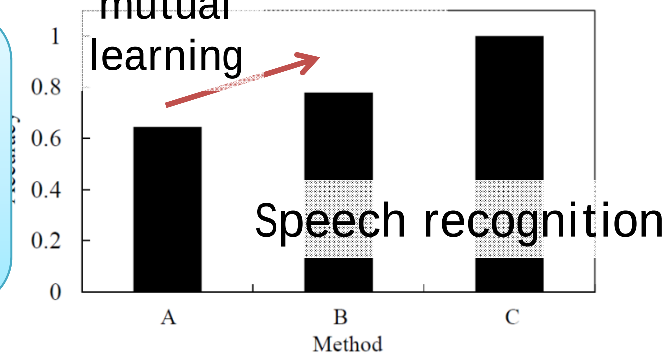
Example sentences used in the experiments



(a) (b) (c)

Obtaining multimodal sensory information



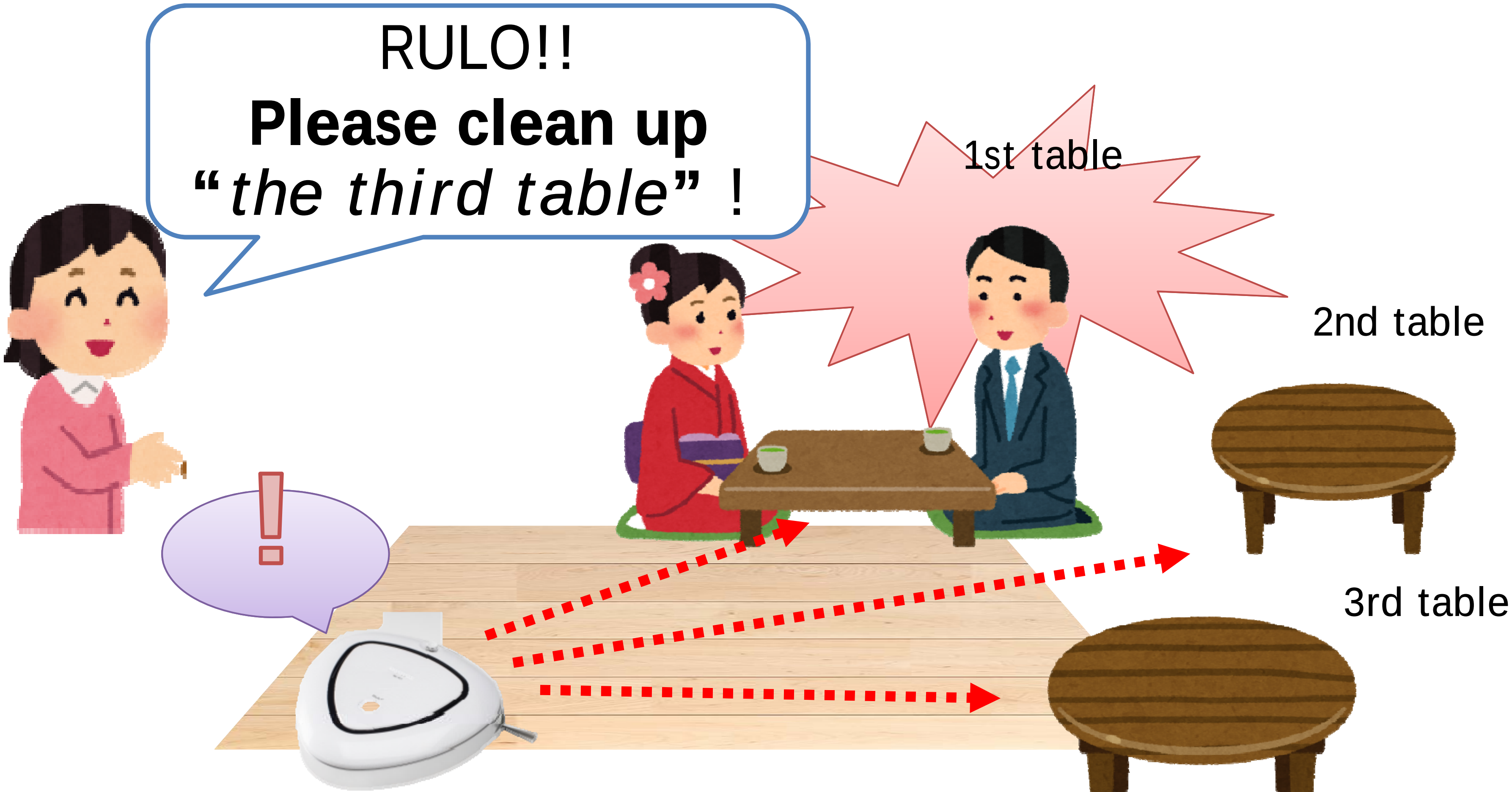


Both **object categorization** and **speech recognition performances** increased using co-occurrence cues.

Contents

1. Introduction
2. Lexical acquisition tasks
    A) Direct phoneme and word discovery from speech signals
    B) Simultaneous acquisition of word units and multimodal categories
    C) Online spatial concept acquisition
3. Future challenges

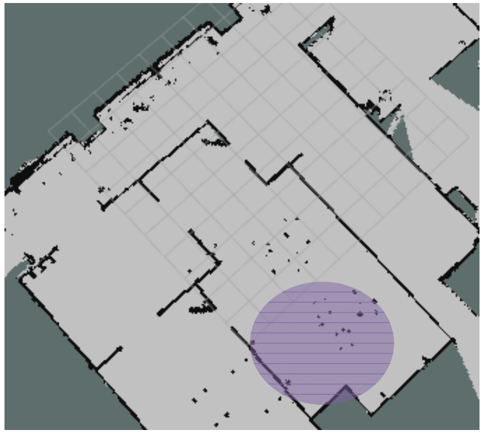
Learning place name is very important for home service robot



*RULO is a cleaning robot produced by Panasonic

Spatial concept is multimodal

Position



Where is the space?

Word

“This is *the third table*”
“*A meeting space*”
“*Under the air conditioner*”

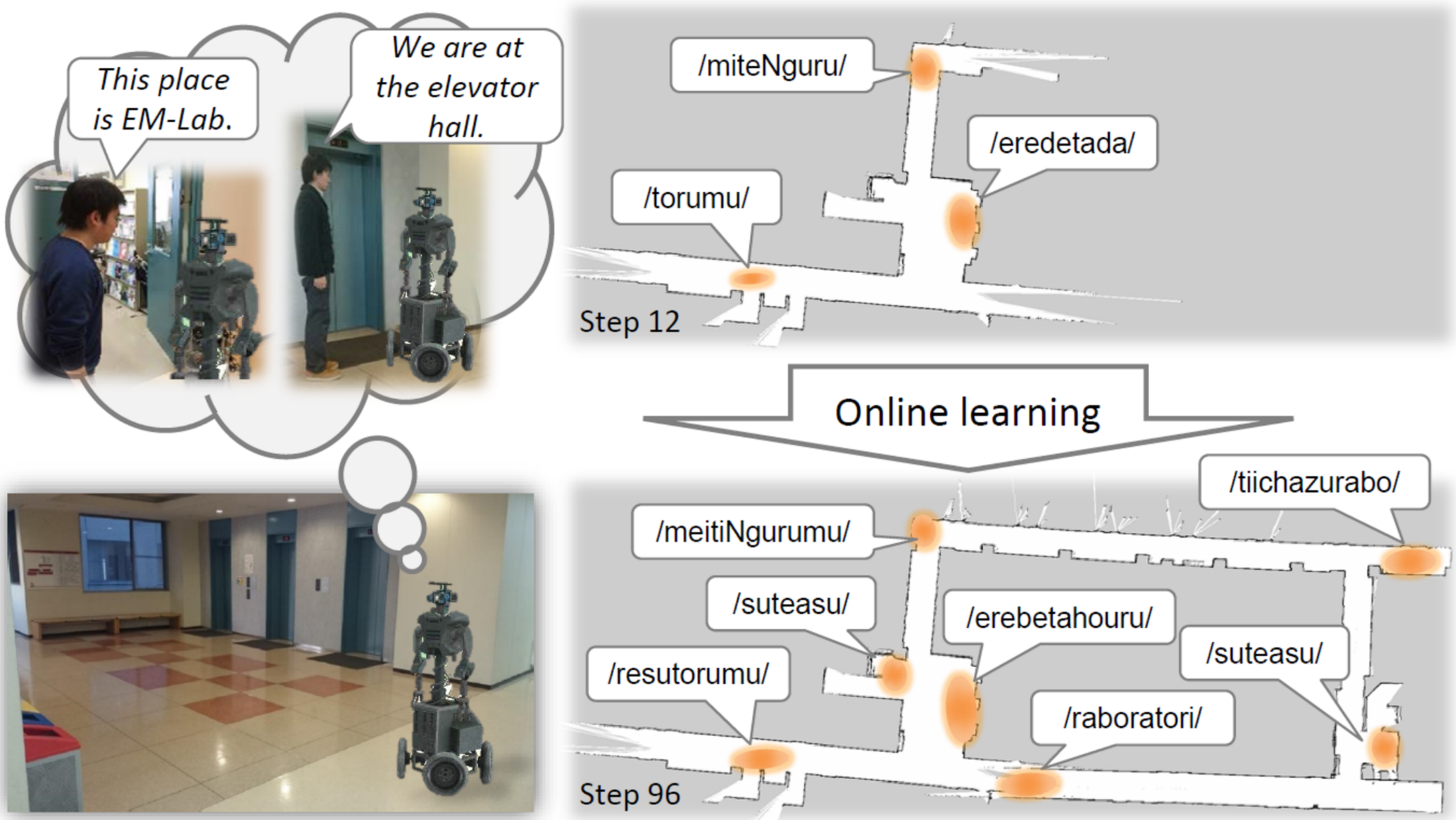
How do they call the space?

Visual information

How does the space look like>

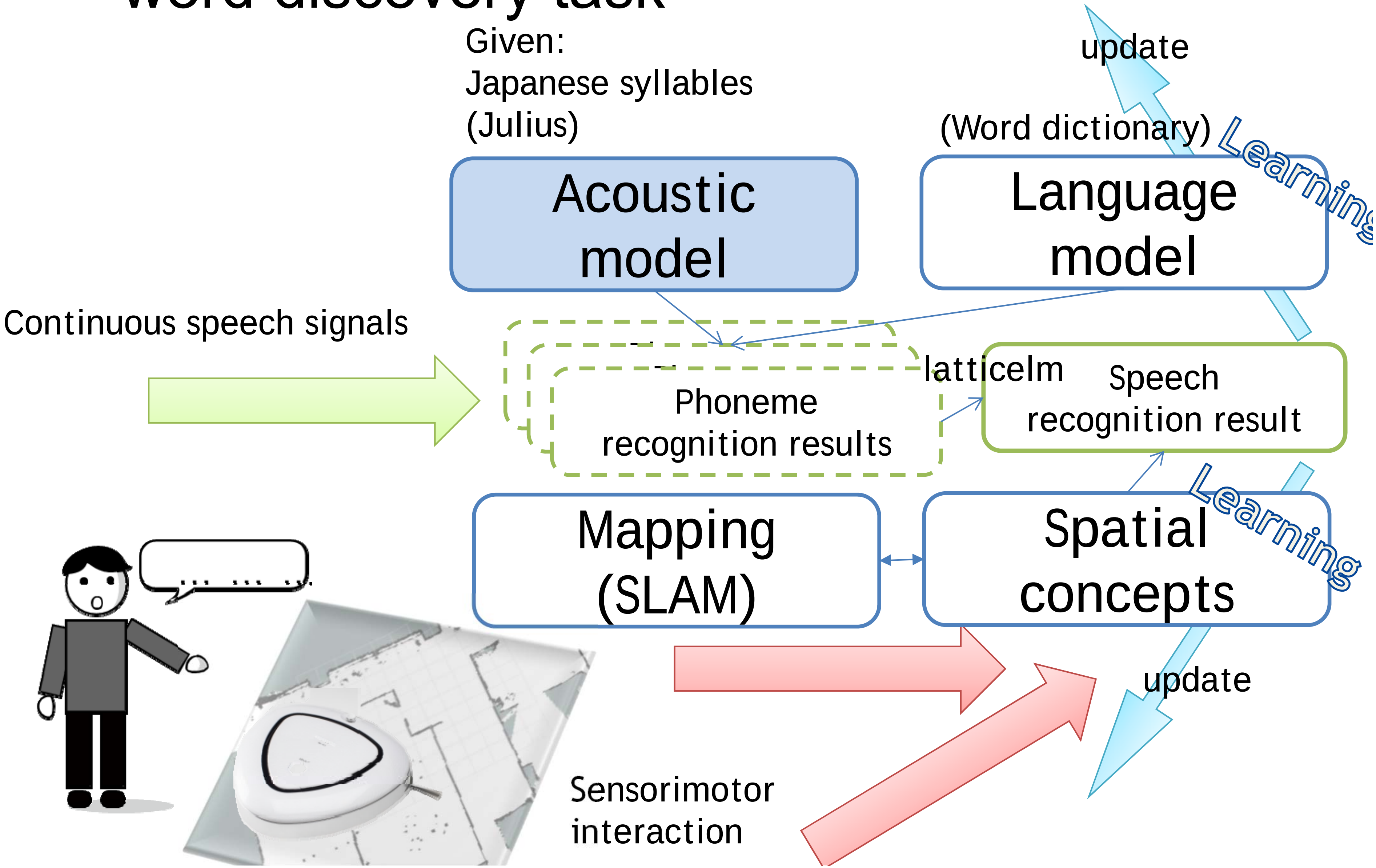
Sound, smell, ,,,

Online spatial concept acquisition method **SpCoSLAM** [Taniguchi+ 2017] (including word discovery task)

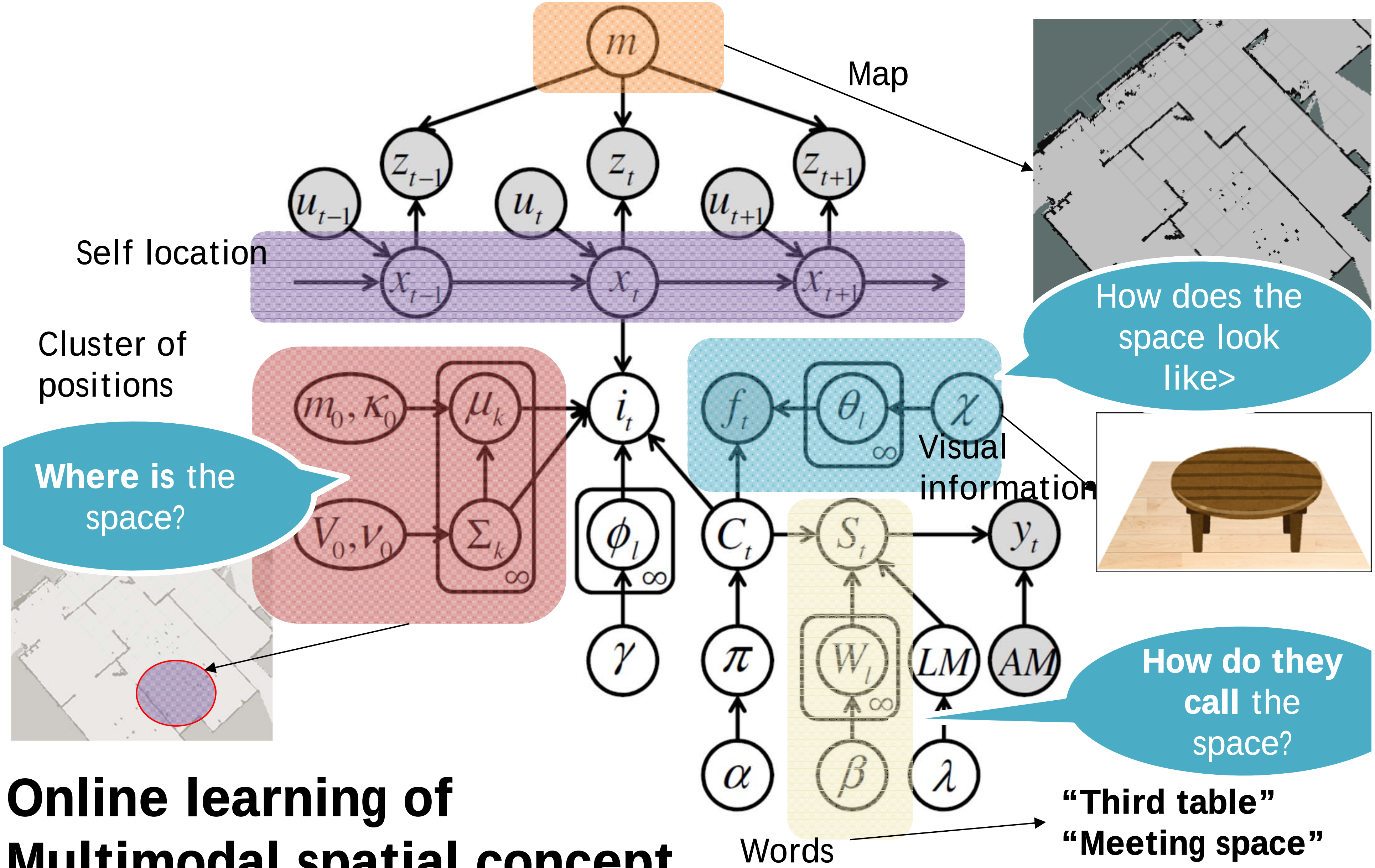


Akira Taniguchi, Yoshinobu Hagiwara, Tadahiro Taniguchi and Tetsunari Inamura, Online Spatial Concept and Lexical Acquisition with Simultaneous Localization and Mapping, IEEE IROS 2017 (submitted)

Online spatial concept acquisition with word discovery task

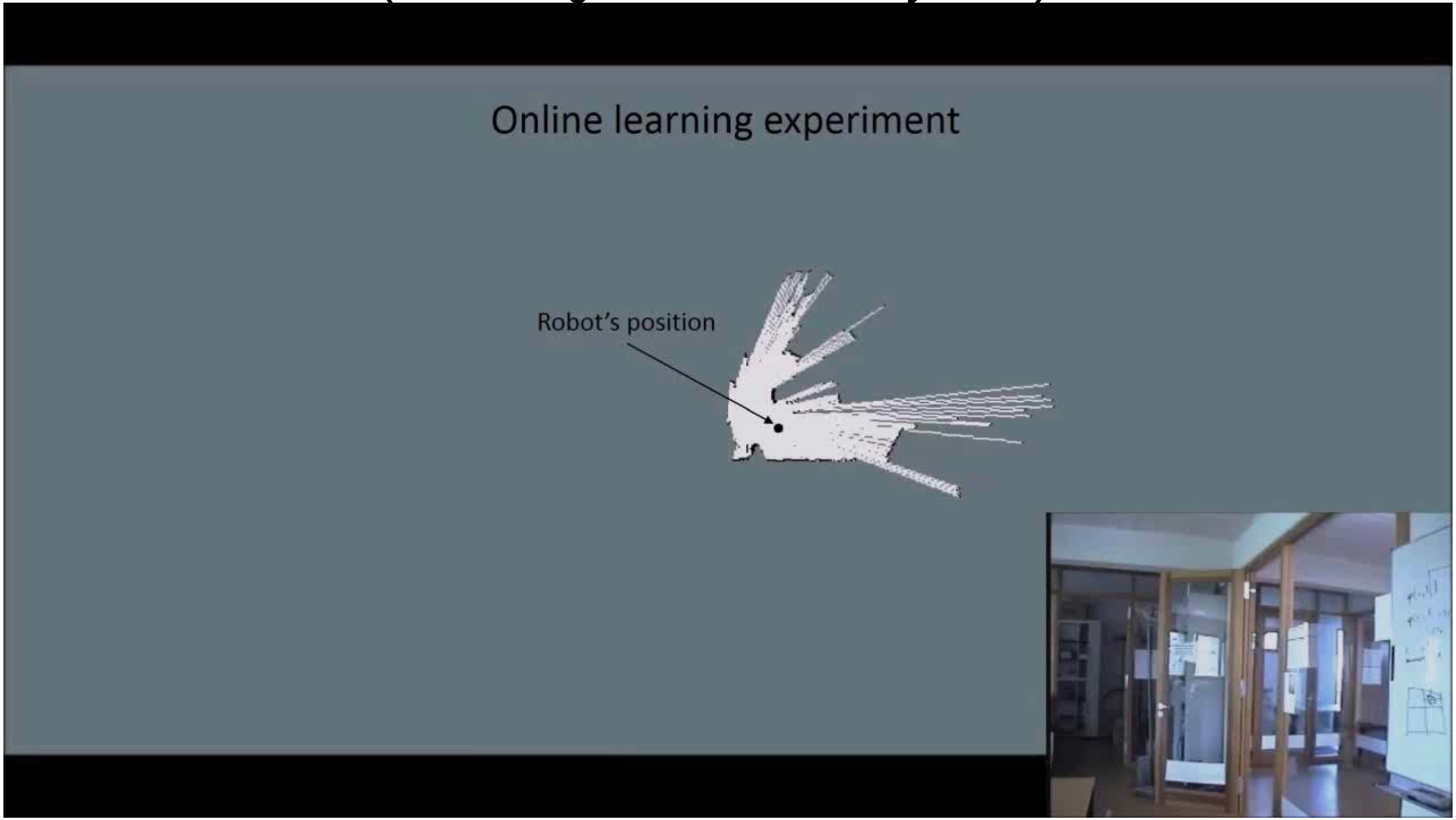


Graphical model of SpCoSLAM



Online learning of Multimodal spatial concept

Online spatial concept acquisition method
SpCoSLAM [Taniguchi+ 2017]
(including word discovery task)

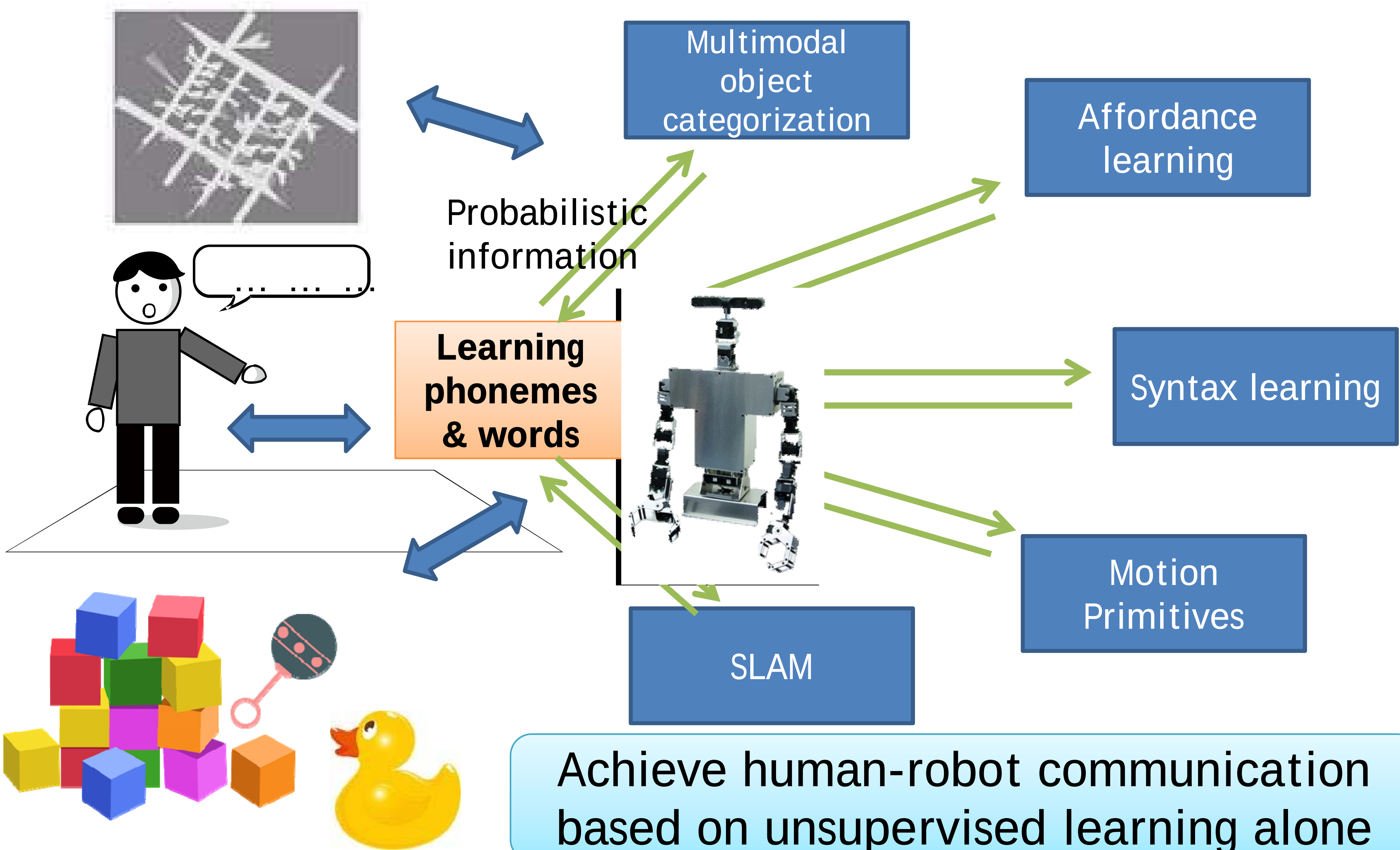


Akira Taniguchi, Yoshinobu Hagiwara, Tadahiro Taniguchi and Tetsunari Inamura, Online Spatial Concept and Lexical Acquisition with Simultaneous Localization and Mapping, IEEE IROS 2017 (submitted)

Contents

Current challenge

Unsupervised learning of lexicons grounded via a robot's sensorimotor information



Achieve human-robot communication based on unsupervised learning alone

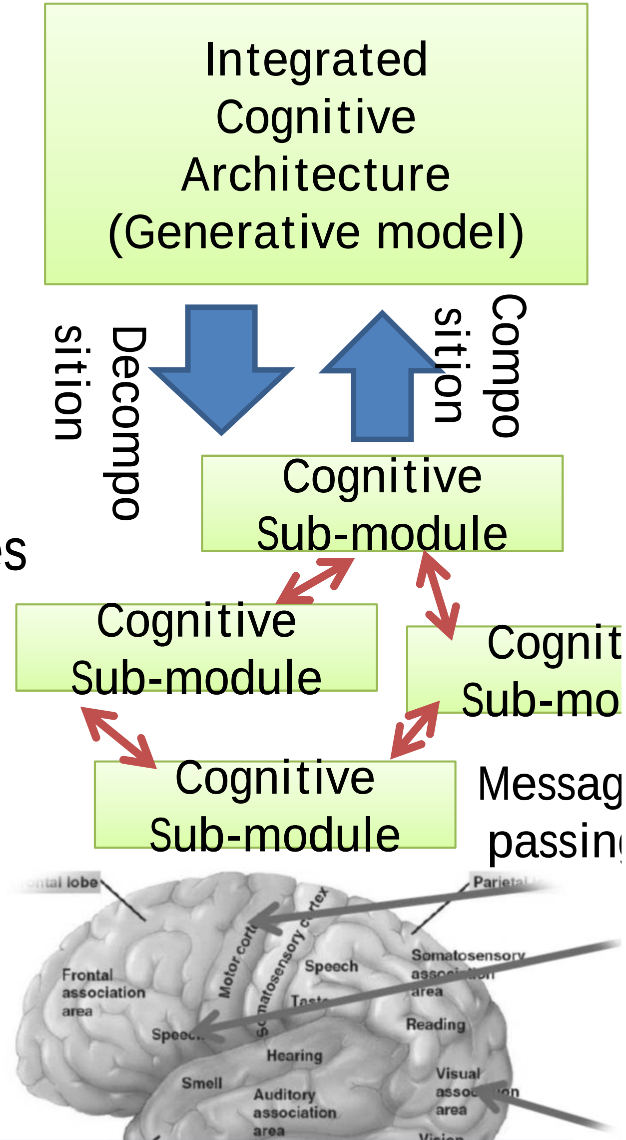
Employing **deep learning** for unsupervised machine learning for symbol emergence in robotics

- **Complex emission distribution** and **automatic feature extraction**
  - ✓Structured Variational Auto Encoding(e.g., GMM+VAE [Johnson 2016])
- Using deep learning in **inference procedure** of probabilistic generative models
  - ✓Amortized inference
- Modeling **language** using recurrent neural network
  - ✓LSTM, GRU, and so on.

Bayesian deep learning

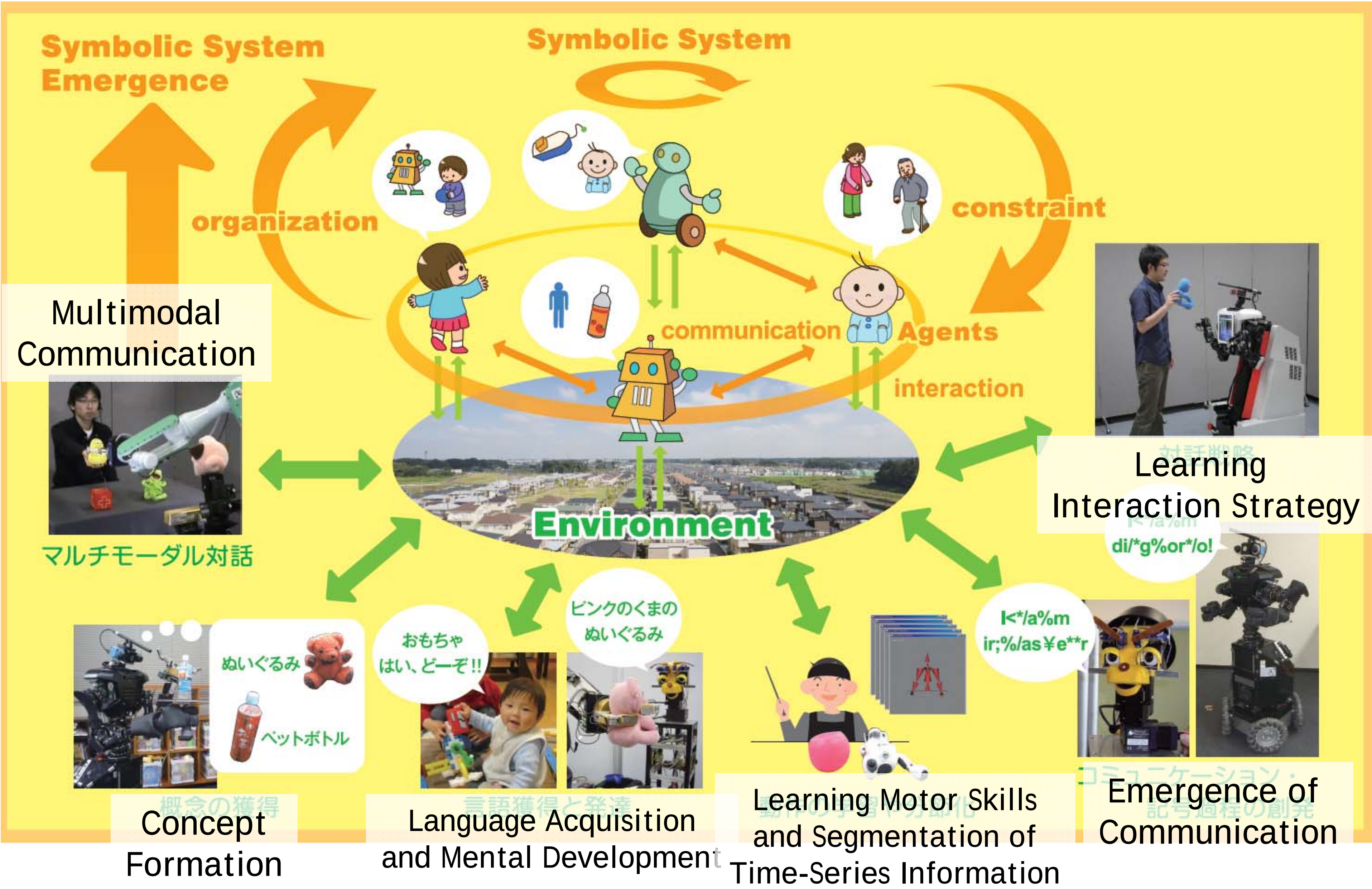
Scaling up.......

- **Software engineering problems: to deal with complex (huge) graphical models representing mutually dependent cognitive modules.**
  - ✓ Developing independent cognitive modules as probabilistic generative models and integrate them into an integrated cognitive module.
  - ✓ Using probabilistic programming. (e.g. Anglican, Venture, Edward)
  - ✓ Employing automatic gradient (e.g. Tensorflow, Autograd) for variational inference



Develop a **developmental cognitive model** by referring to **human whole-brain architecture**

Symbol Emergence System



Information

Special Thanks

- Ritsumeikan University
  - R. Nakashima, S. Nagasaka, A. Taniguchi,
- UEC
  - T. Nagai, T. Nakamura, T. Araki, Y. Ando
- Okayama Pref. Univ.
  - N. Iwahashi

Email: taniguchi@ci.ristumei.ac.jp
Twitter: @tanichu
Facebook: Tadahiro Taniguchi
HP: http://www.tanichu.com

Funded by

科研費
KAKENHI

JST CREST

The 2nd Workshop on Machine Learning Methods for High-Level Cognitive Capabilities in Robotics 2017 @IROS2017