Intersensory Causality Modeling using Deep Neural Networks

Kuniaki Noda, Hiroaki Arie, Yuki Suga and Tetsuya Ogata

Waseda University, Tokyo, Japan

October 15, 2013
SMC2013 @ Manchester, UK

Multimodal integration

Humans enhance perceptual precision and reduce ambiguity by integrating multimodal information (Ernst2004, Stein1993)

Computational models that replicate human multimodal integration ability may contribute to robots working in human environments

Action-effect causality understanding

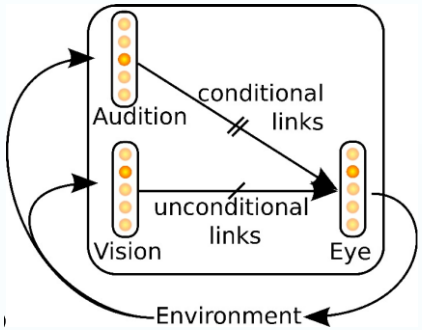
Sensory prediction in response to the self action

Construct abstracted internal representation for recognition

Multimodal integration learning

- Applications for the learning of robotic behaviors

(Sauser and Billard, 2006)

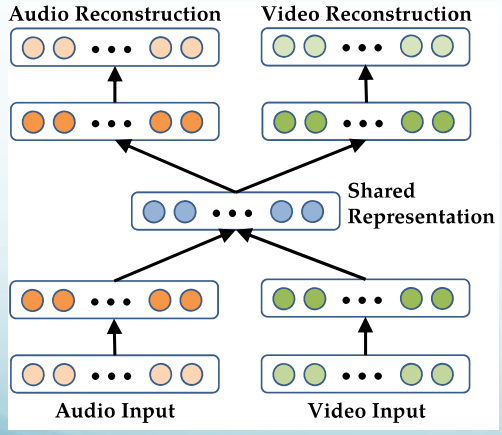


(Pitti et al., 2012)

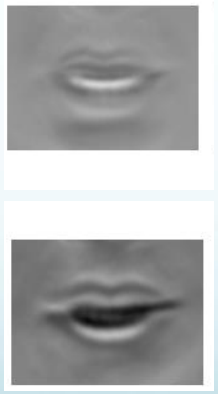
Limitation in scalability

Dedicated sensory feature extraction mechanisms

- Multimodal integration by deep neural networks



(b) Bimodal Deep Autoencoder



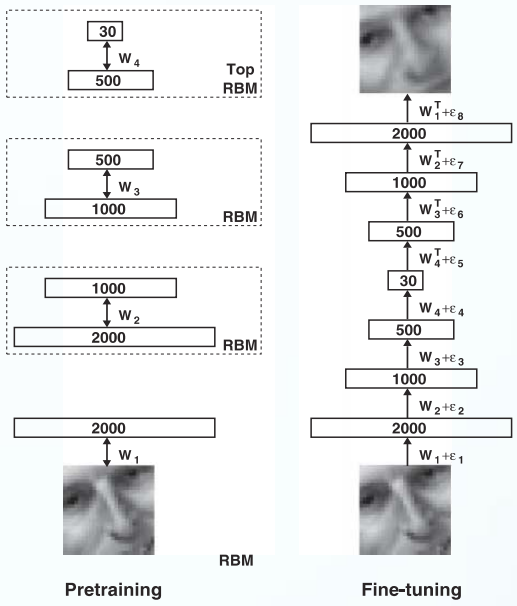
(Ngiam et al., 2011)

Scalable, but static modalities

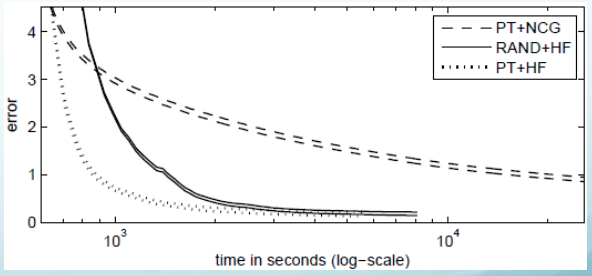
Sound, image, text, etc.

Deep Neural Networks

- *G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, 2006.*
  - Epoch-making article which leads to the current trends for the deep learning
  - Utilize RBM for training single layer network in the pre-training phase, followed by the entire layer training in the fine-tuning phase
- *J. Martens, "Deep learning via Hessian-free optimization," ICML, 2010.*
  - Utilize quadratic programming
  - Pre-training is not required
  - Optimization algorithm based on the Newton's method contributes in faster convergence



(Hinton, 2006)

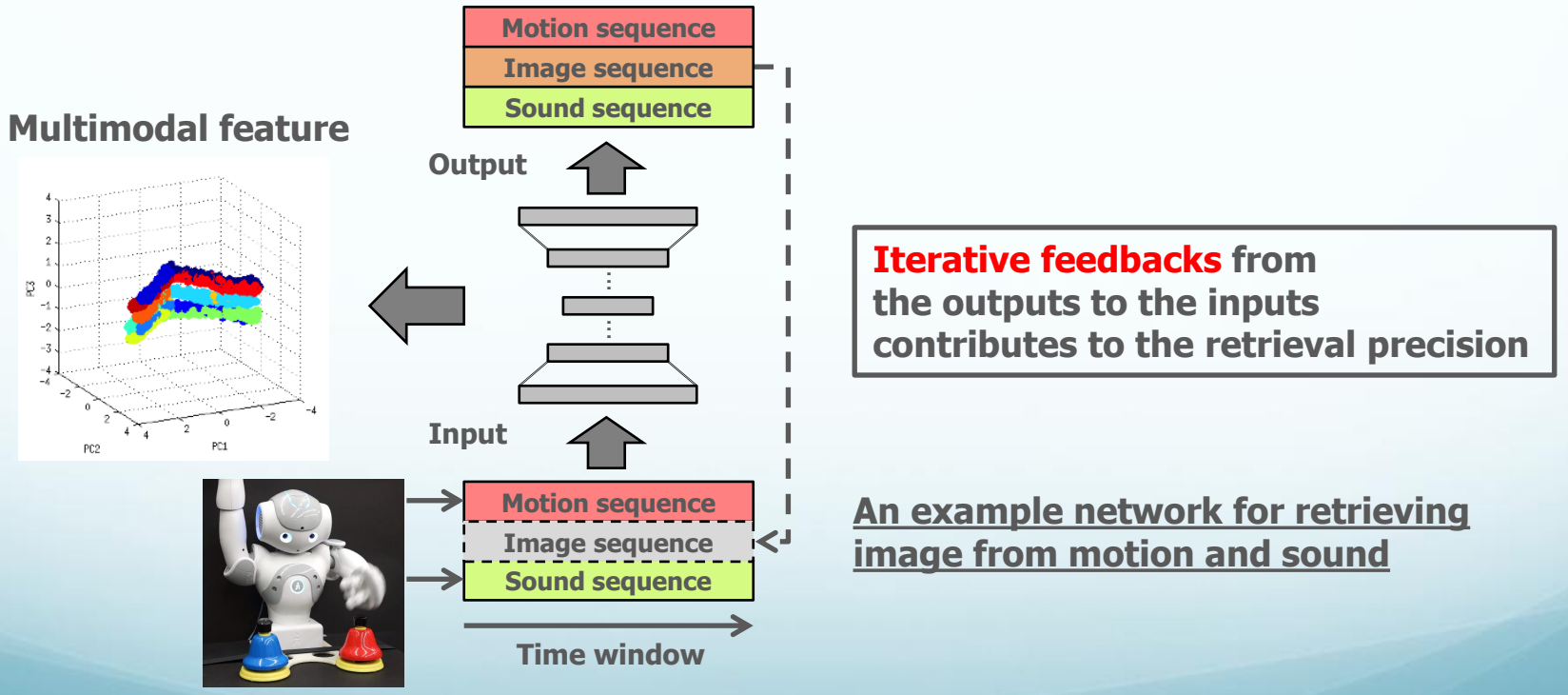


(Martens, 2010)

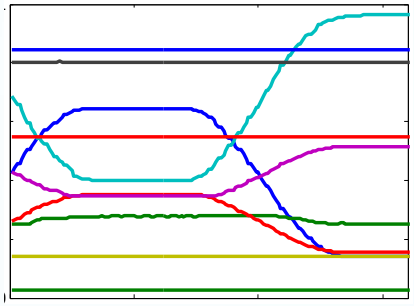
We adopt Hessian-free optimization for the training algorithm

Time-delay autoencoder

- Temporal sequence learning by a deep autoencoder
  - Compresses multimodal temporal segments
  - Models inter-dimensional correlations
  - Retrieves temporal sequence in cross-modal



Bell ring task by a humanoid robot



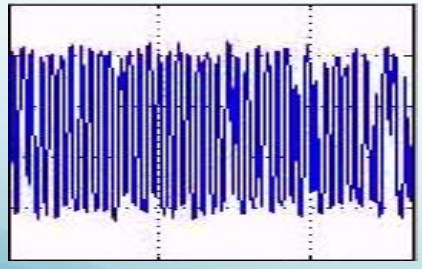
Joint angles

Motion

Audio

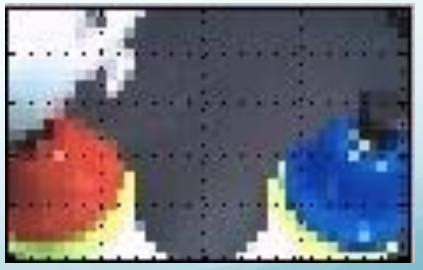
Visual

Bell ring task



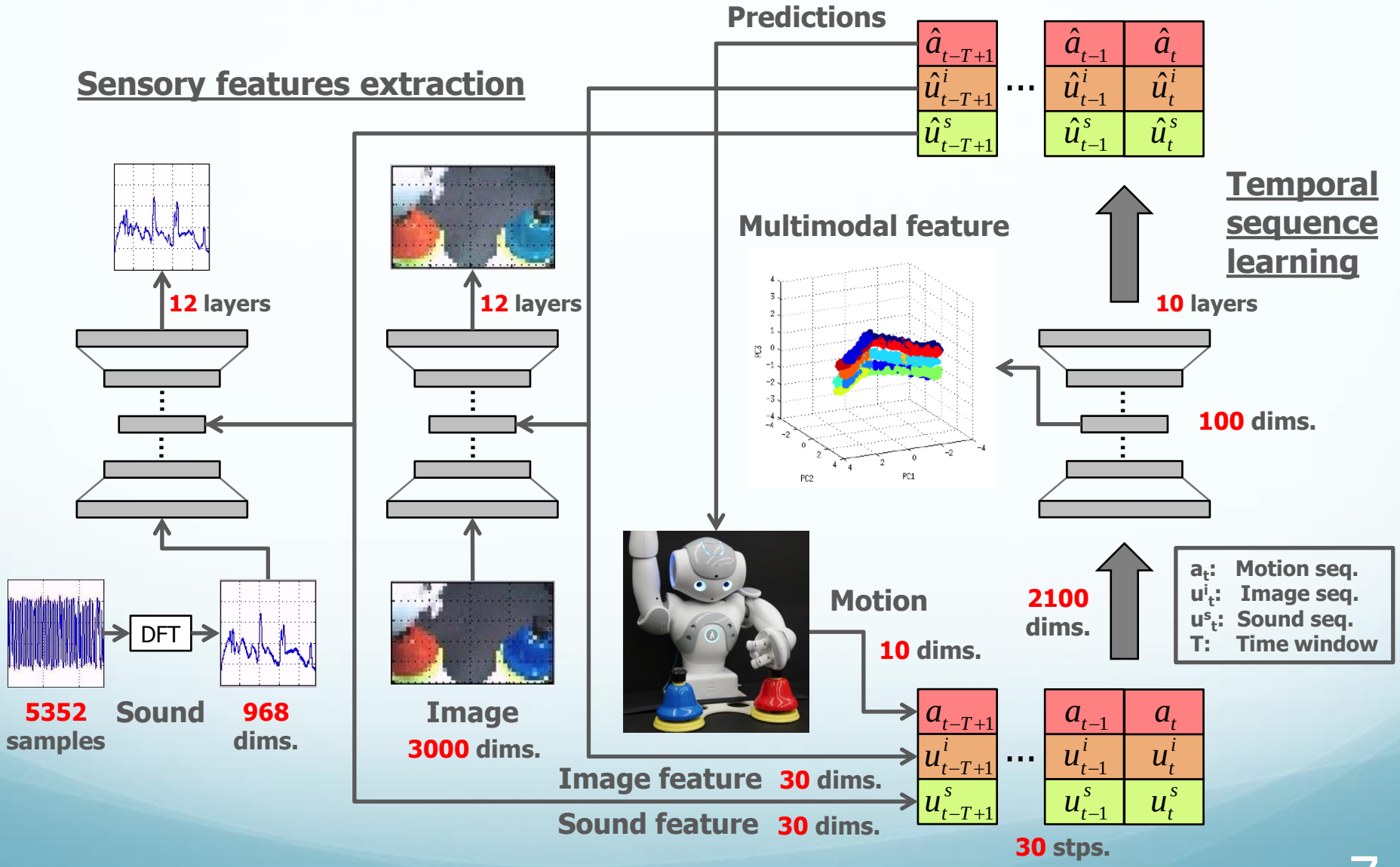
Sound waveform

Understanding
the intersensory consequences
of modality-specific stimuli



Color image

Multimodal integration mechanism

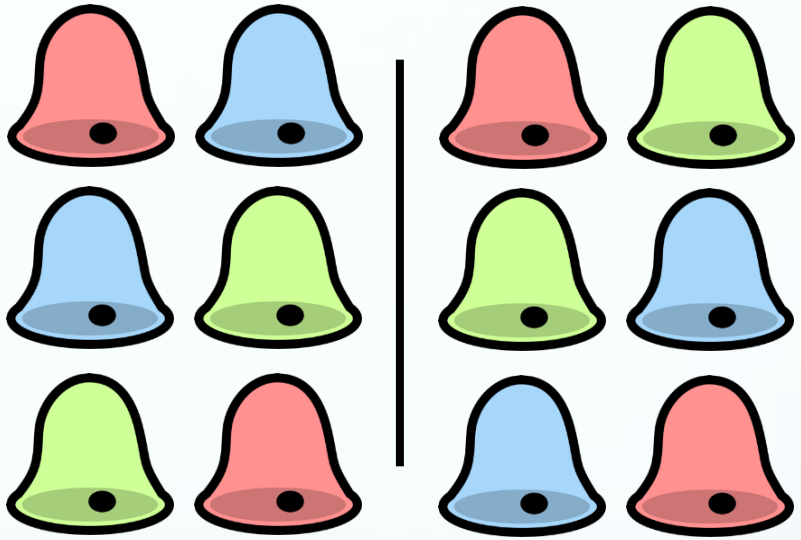


Bell placement configurations

3 kinds of bells

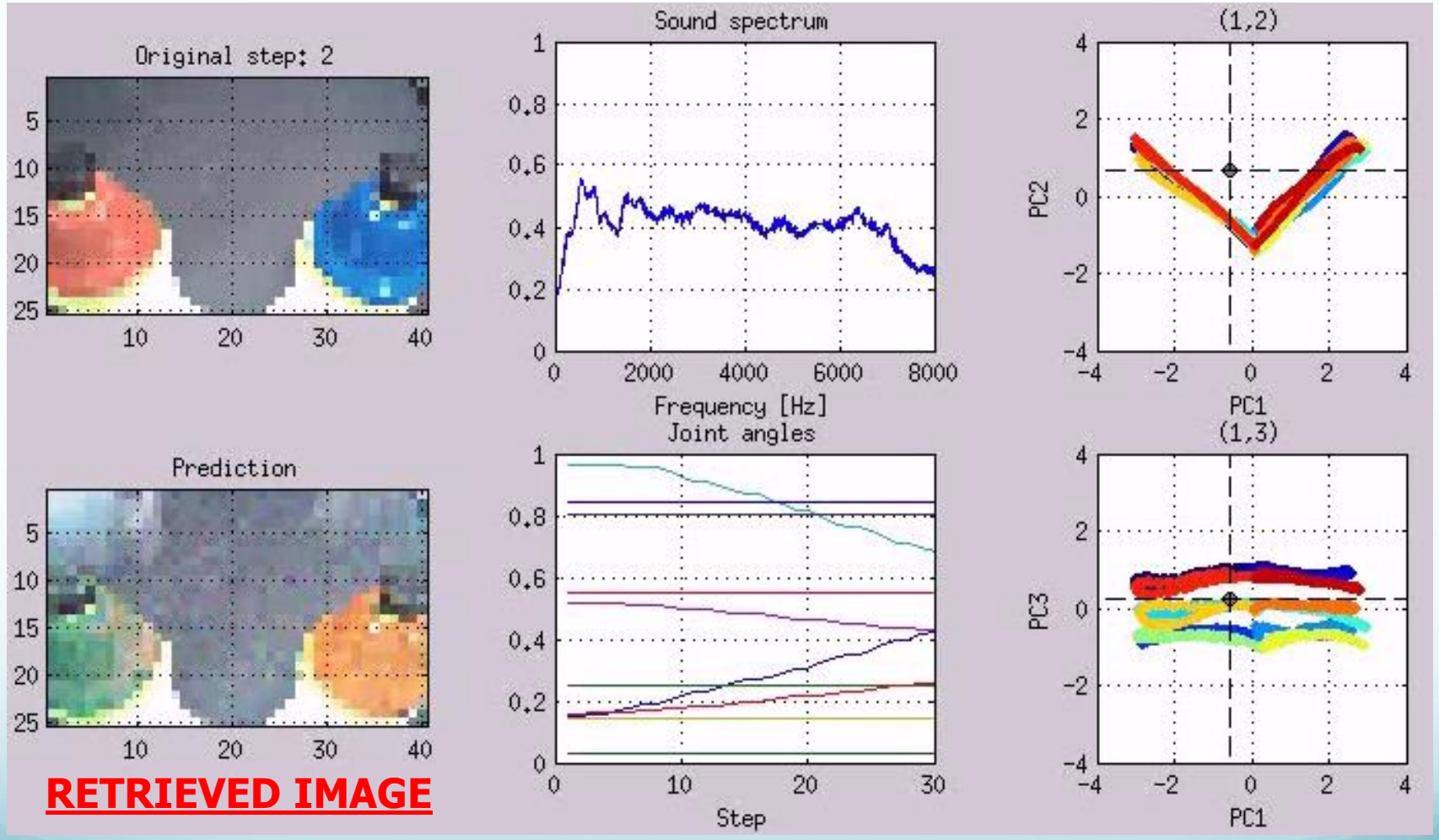
Color	Pitch notation
RED	C
GREEN	F
BLUE	A

6 placement variations

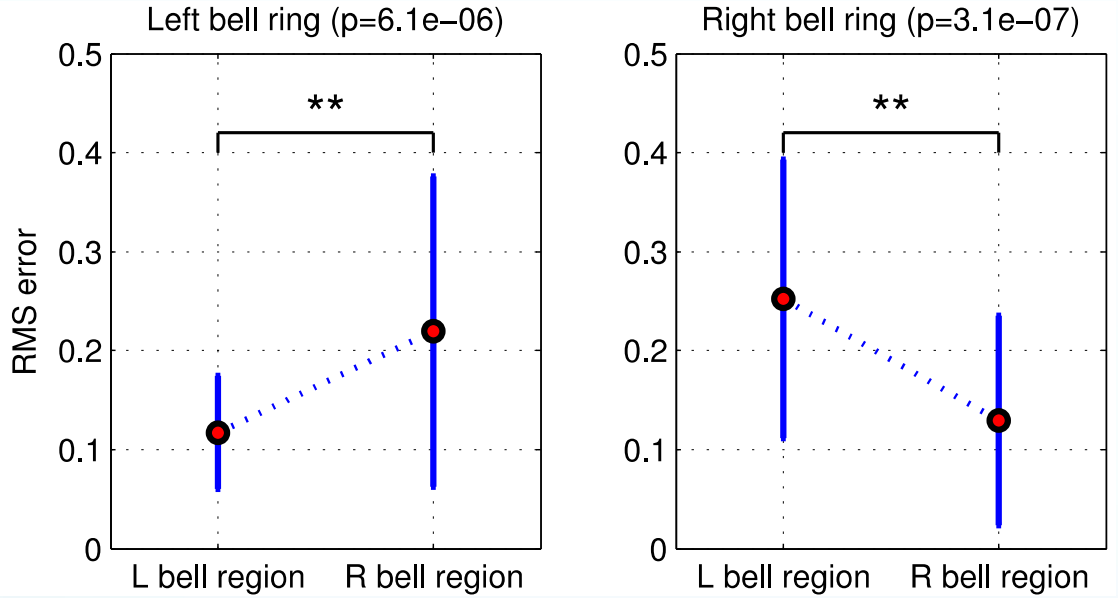


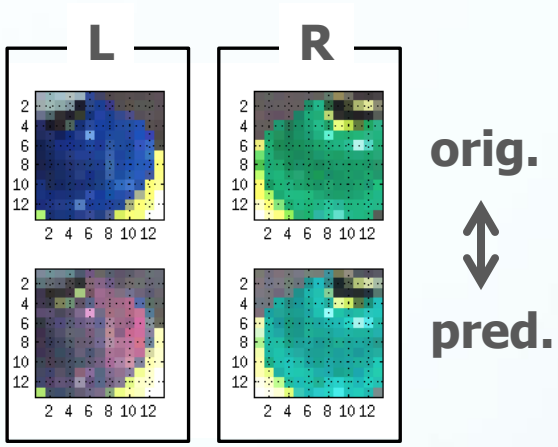
Hit either one of the two

Image retrieval from sound and motion



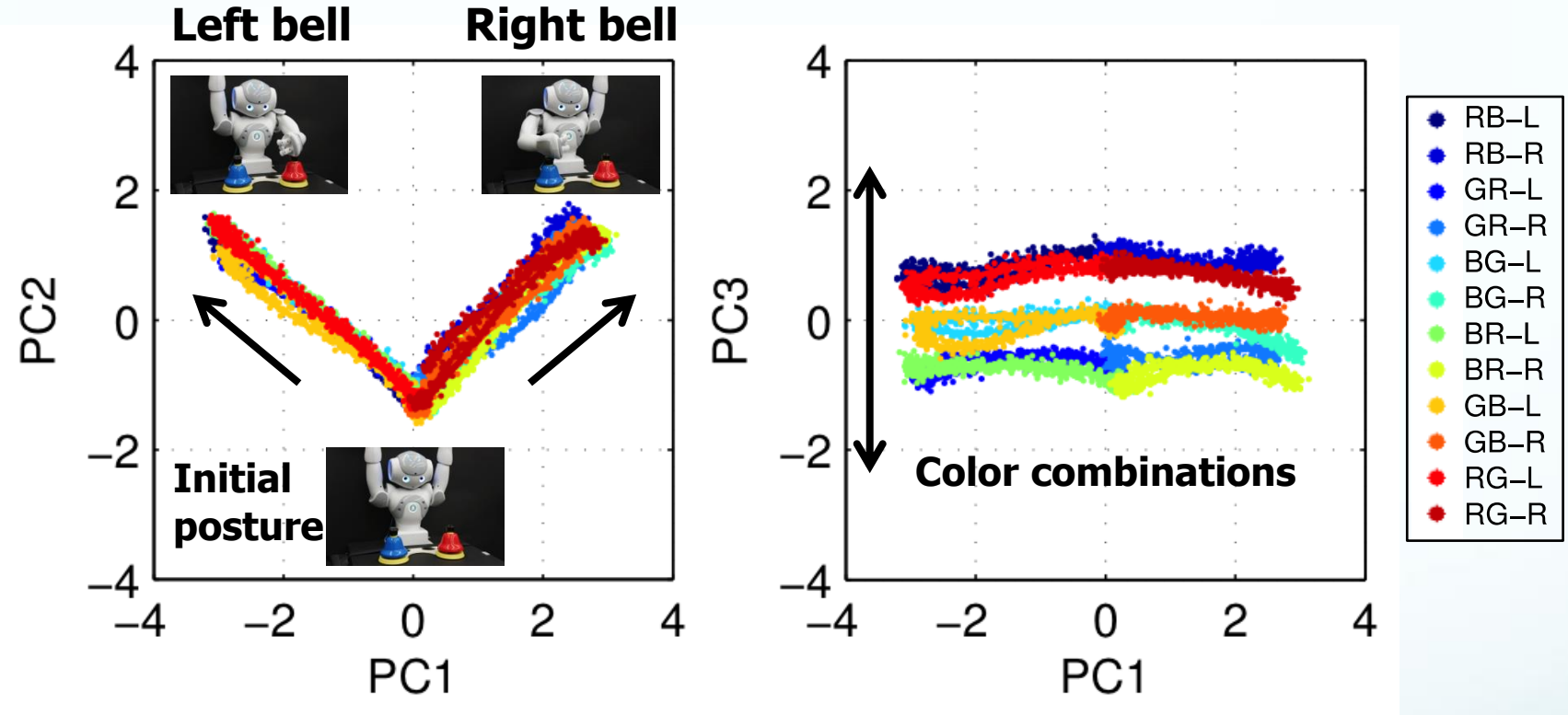
Correlation between generated motion and retrieved images





Bell image prediction accuracy for the region where the arm motion coincide outperforms the other

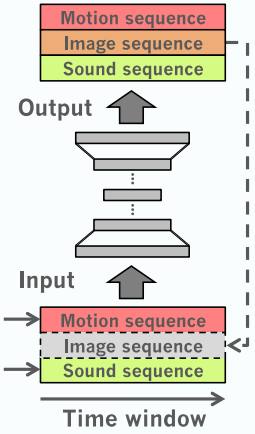
Sensory-motor integrated feature space



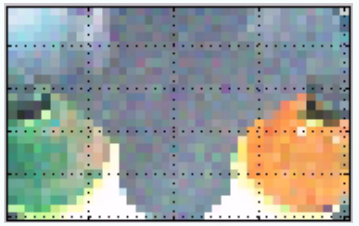
Phase-wise motion is represented on a plane and the bell placement configuration is structured on the third axis

Conclusion

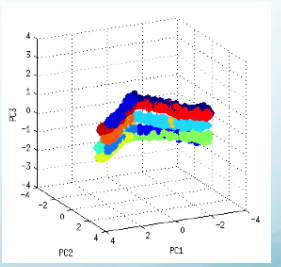
- Multimodal sensory-motor integration learning
  - A novel computational framework for the temporal sequence learning utilizing a deep neural network



- Intersensory causality modeling
  - Image sequence retrieval based on the the acquired sensory-motor causality



- Self-organization of multimodal feature space
  - Phase-wise sensory-motor integrated feature is structured on the modal dependent coordinates



Thank you!

The authors would like to thank **the Hara Research Foundation** for their financial support.

The work has been supported by JST PRESTO "Information Environment and Humans" and MEXT Grant-in-Aid for Scientific Research on Innovative Areas "Constructive Developmental Science" (24119003).