Mathematical modeling of binding semantic memory with visual information

Hideyuki Kato
Faculty of Science and Technology,
Oita University
700 Dannoharu, Oita,
Oita 870–1192, Japan
0000-0002-9447-5677

Haruki Terakawa
Graduate School of Engineering,
Oita University
700 Dannoharu, Oita,
Oita 870–1192, Japan
0009-0001-5509-5786

Yoshihiro Yonemura
School of Systems Information Science,
Future University Hakodate
116–2 Kamedanakano-cho, Hakodate,
Hokkaido 041–8655, Japan
0000-0002-7527-233X

Yuichi Katori
School of Systems Information Science,
Future University Hakodate
116–2 Kamedanakano-cho, Hakodate,
Hokkaido 041–8655, Japan
0000-0003-2773-0786

***Abstract*—This study investigates visual information processing and its integration with semantic memory with a hierarchical predictive coding with reservoir computing model. The model explores how visual elements like color and shape interact with semantic memory, focusing on the reverse Stroop effect. The model processes images of gendered toilet symbols, demonstrating that East Asian models, which use color cues for gender identification, exhibit quicker recognition than other models. The perception is accurate for images with consistent color and shape information but inaccurate for images with conflicting color and shape information. The results illustrate that the model effectively simulates the reverse Stroop effect, revealing insights into visual and semantic memory interactions.**

***Index Terms*—reservoir computing, predictive coding, modeling cognitive process**

I. Introduction

Animals process sensory information from the external environment and recognize it. There are five types of sensory information: touch, taste, smell, hearing, and vision [1]. In the case of human beings, visual information dominates more than 80% of all the sensory information for recognition [2], which indicates that visual information is one of the most critical factors in recognition. Recognition requires integrating and unifying the sensory information with memories stored in the medial temporal lobe [3]. The viewpoint of contents classifies memory into declarative and non-declarative memories [1]. Episodic and semantic memories belong to the former, while the latter includes procedural memory and other types of memories. Investigating the mechanisms of sensory information processing and binding it with memories is meaningful to understanding the recognition process.

Reservoir computing (RC) is a computational framework representing approximately dynamical systems with a fixed reservoir and plastic readout [4]. Compared with a deep learning framework, RC only modifies the weights of readout connections, so its computational costs in learning are much lower. Additionally, a reservoir does not need to be adaptively tuned differently from deep neural networks, and then several physical systems are applicable as a reservoir [5]. Due to such advantages, RC is paid much attention to and is widely studied. Some studies employ RC as a model of the brain. For instance, RC can generate multiple patterns via the first-order reduced and controlled error (FORCE) learning algorithm [6]. A multi-modal model consisting of predictive coding with RC (PCRC) accounts for visual and auditory information and their bindings through the task of associating corresponding handwritten images from spoken digits [7].

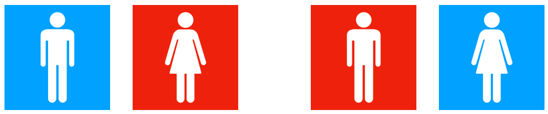
Following the above investigations, using a hierarchical PCRC model, this paper will study visual information processing and binding it with semantic memory stored in the medial temporal lobe through the reverse Stroop effect as an example of this process. Therefore, we will focus on the colors and shapes of visual information and the gender concept of semantic memory.

II. Materials and methods

A. Stroop effect and reverse Stroop effect

Stroop, an American psychologist, reported a phenomenon in his psychological experiment [8]. In his experiment, participants should answer the names of colors printed in incongruent ink colors. For instance, the word “blue” was printed with green ink. Then, the participants received an inquiry about the actual color of the word or ink rather than the word written by the characters. As a result, the participants needed to answer the question longer than under the condition where the consistent ink color printed the word. In this phenomenon, called the Stroop effect, the word involving the information interferes with the information about the color. The opposite

JSPS Grants-in-Aid for Scientific Research JP20H00596, JP23K21706, JP21K12050, and JP24H02330. Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University.



(a) Normal images (b) Abnormal images

Fig. 1: Examples of toilet symbols as visual information.

direction of the interference is the reverse Stroop effect [9]. For instance, it takes a longer time to recognize the gender implied in toilet symbols like Fig. 1(b) since toilet symbols like Fig. 1(a) are common or familiar to East Asian people. People in the US and Europe, where there is no use of color for the distinction between males and females, do not exhibit this trend.

B. Hierarchical predictive cording model with reservoir computing

In this paper, we constructed a multimodal model of PCRC to study how visual information is processed in the visual cortex and binds with semantic memory in the medial temporal lobe (Fig. 2) [10]. The network architecture is roughly the basis of the physiological findings [1]. In brief, the model consists of three PCRCs. Two of the three PCRCs, color and shape areas assumed as the thin and pale stripe in V2, independently process colors and shapes in the visual information of images, and the other one, integration area assumed as the higher visual cortices and the medial temporal lobe, integrates and unifies the information from the two downstream PCRCs and the gender concept implied in the visual information and feeds back the processed stuff to the two downstream PCRCs. Neurons in the shape and integration areas are homogeneous, while the color area involves color-selective neurons responding to reddish, greenish, and blueish colors, respectively [11].

The color and shape areas learn the corresponding information in the learning phase. The integration area learns the integrated information of the two downstream areas and the gender concept of the visual information. The FORCE learning algorithm updates the synaptic weights of output connections in all the parts of the PCRCs. After learning, we stopped giving the gender information to the integration area and removed the path of the error feedback to the integration area. The model then predicts the corresponding gender of the visual information only from its colors and shapes in the test phase.

An input to the model is an image with $N \times M$ pixels. The preprocessing decomposes the color and shape elements from the image. The pixel values in the RGB coordinate map to the CIE coordinate [12], and then its origin shifts to the white. The polar coordinate format of the color of each pixel in the shifted CIE coordinate allows us to express the color and its strength with the phase and amplitude. The phase ranges in $[\theta_{\mathrm{M}}, \theta_{\mathrm{O}})$, $[\theta_{\mathrm{O}}, \theta_{\mathrm{C}})$, $[\theta_{\mathrm{C}}, \theta_{\mathrm{M}})$, where θ_{M}, θ_{O}, and θ_{C} are the phases of Magenta, Orive, and Cyan, correspond to reddish, greenish, and blueish, respectively. The three colors in the RGB space locate at $(0.5, 0, 0.5)$, $(0.5, 0.5, 0)$, and

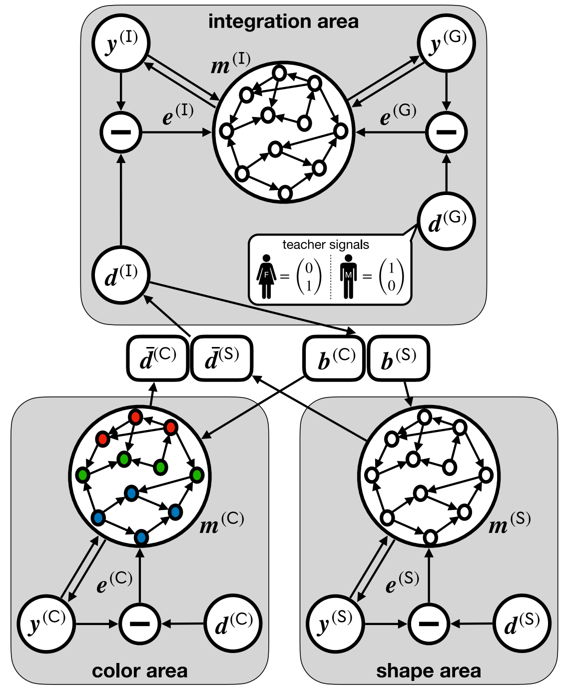


Fig. 2: Schematic of PCRC model.

$(0, 0.5, 0.5)$, respectively. We assumed the total values of the amplitudes in every phase range as the input to the color-selective neurons. The shape element of the gray-scaled input image filtered by the Gabor filter [13] is flattened and transformed to a vector with NM dimension. The Gabor filterings with the angles every $\pi/6$ from 0 to $5\pi/6$ radians acting as direction-selective neurons result in the six filtered images. Each pixel is the mean value of three images randomly selected from the six filtered images before being flattened. Aligning all the vectors for the images makes a matrix with $NM \times n$ where n is the number of all the images given to the model in the learning phase. t-SNE reduces the dimension of the matrix to $3 \times n$, where each column in the matrix is a vector corresponding to each image [14]. We applied the vector as an input to the neurons in the shape area. An input to the neurons in the integration area is a 2-dimensional one-hot vector indicating the gender: $(1, 0)^{\top}$ and $(0, 1)^{\top}$ for the male and female, respectively, in addition to the input from the two downstream areas.

C. Input data sets

To investigate visual information processing and its binding with semantic memory through the reverse Stroop effect, we collected 100 images of toilet symbols illustrating a pictogram of the male or female on the internet. Half of them were male, and the others were female. The design of the collected images had the typically observed colorings in some East Asian countries, especially Japan: blueish for the male and reddish for the female (Fig. 1(a)). The corresponding color to the gender filled the inside or outside of the pictograms, and the other side was white. We called the images the normals.

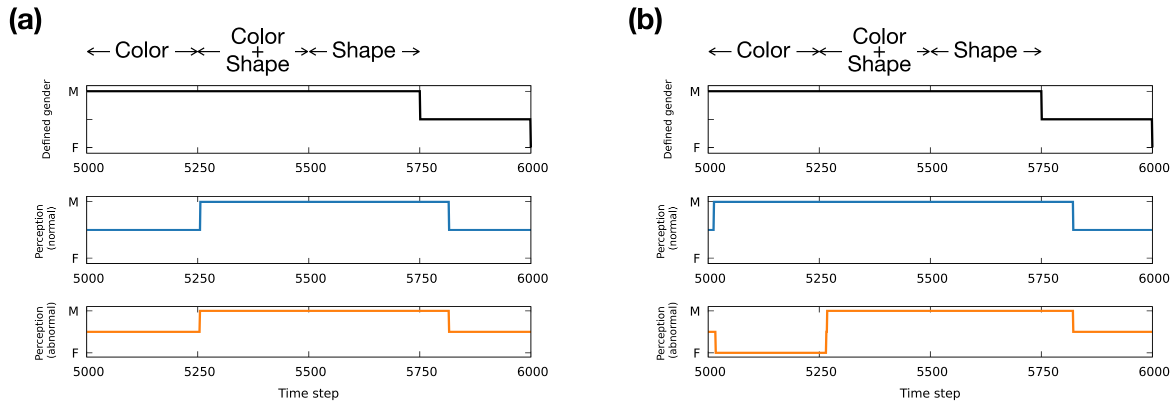


Fig. 3: An example of the gender perception in the hierarchical PCRC model to an image of a male pictogram. (a) The Western and (b) the East Asian types. In each panel, the defined gender of a given image, which corresponds to the shape indicating gender (upper), the model perceptions for a normal image (middle) and for an abnormal image (bottom). The presentation of an image is for the first 500 steps. Because processing in the shape area delays 250 steps, the integration area only process the information from the color area in the first 250 steps, the information from both color and shape areas in the second 250 steps, and then only the information from the shape area in the third 250 steps.

From the normals, we created the abnormals by exchanging the color corresponding to the gender: blueish for the female and reddish for the male (Fig. 1(b)). In addition, the binarization of the normals allowed us to obtain black-and-white toilet symbols.

D. Protocols in learning and test phases

We examined two types of models: the Western and East Asian types. In the learning phase, the inputs to these two types of models are different. Learning the different inputs reflects differences in environments or cultures between the countries. Inputs to the Western type in the learning phase are the black-and-white images of toilet symbols since there is no use of color to distinguish males and females. In contrast, the East Asian type learns the normals like Fig. 1(a) that are common or familiar in these countries. The FORCE learning independently updates the output connections in the color and shape areas to reduce the errors between the preprocessed inputs and the outputs in both model types. After this independent learning of the two downstream areas, the integration area learns the inputs from the two downstream areas and the gender of the visual information. Because of the study of the reverse Stroop effect, we defined the gender corresponding to the shape information of the pictograms represented by the one-hot vector as a teacher signal. In a learning protocol, the integration area learns the color and the gender, and then the shape and the gender.

In the test phase, we presented the normal and abnormal images to the model and observed the performance of the gender prediction. The processing in the shape area follows the one in the color area [15]. The shape area starts processing 250-steps after the color area from the physiological insight. Since the model outputs a vector $\boldsymbol{y}^{(\mathrm{G})} \in \mathbb{R}^2$ as the predicted gender of images, we defined the male for $y_1^{(\mathrm{G})} > y_2^{(\mathrm{G})}$ and the female otherwise.

III. Results

The Western type exhibits invariant perception for an image presentation (Fig. 3(a)). The model perception matches perfectly with the defined gender of the given image. Additionally, the model perception does not depend on whether a given image is normal or abnormal, which indicates that color information in an image does not affect the gender perception of a given image. Then, the Western type of the model only utilizes the shape information for the gender perception because the given images for its learning are black-and-white.

The perception in the East Asian type is consistent with that in the Western type for the other countries for the normal (Fig. 3(b) middle). The East Asian type gets the gender perception about 240-steps earlier than the other country model (Fig. 3(a) and (b) middle). This is due to the use of the color information of the given image for the perception. Contrarily, in the abnormal case, the model perception mismatches the defined gender of the presented image at the beginning of the image display (Fig. 3(b) bottom). Processing only color information causes the misleading. However, processing shape information contributes to recovering the gender perception and brings it to the correct gender. The East Asian type takes a long time to correctly recognize the gender of a given image so that the East Asian type reproduces the reverse Stroop effect. These facts indicate that the East Asian type of the model recognizes the gender implied by the visual information with both color and shape elements in the visual information.

IV. Conclusions

In this study, we proposed a multi-modal predictive coding model with reservoir computing to study bindings of semantic memory with visual information via the reverse Stroop effect. As a result, the East Asian model appropriately recognizes the genders of given images when the gender indicated by colors confirms that of shapes in images. In contrast, the conflict between the gender of colors and shapes in images forces the model to recognize the correct gender more times. However, we have not observed the behaviors in the model for people in other countries. These facts suggest that our proposal quantitatively models visual information processing and binding it with semantic memory. In other words, the model successfully mimics the reverse Stroop effect.

REFERENCES

[1] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience: Exploring the brain: Fourth edition*, 2015.

[2] The Illuminating Engineering Institute of Japan, Ed., *The Guide to Indoor Lighting (in Japanese)*. Denkishoin, 1978.

[3] K. Sakai, "Neuronal organization of cognitive memory," *Biophysics (in Japanese)*, vol. 34, no. 5, pp. 179–184, 1994.

[4] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Computation*, vol. 14, pp. 2531–2560, 2002.

[5] K. Nakajima, "Physical reservoir computing—an introductory perspective," *Japanese Journal of Applied Physics*, vol. 59, 2020.

[6] D. Sussillo and L. F. Abbott, "Generating coherent patterns of activity from chaotic neural networks," *Neuron*, vol. 63, pp. 544–557, 2009.

[7] Y. Yonemura and Y. Katori, "Network model of predictive coding based on reservoir computingm for multi-modal processing of visual and auditory signals," *Nonlinear Theory and Its Applications, IEICE*, vol. 12, pp. 143–156, 2021.

[8] J. R. Stroop, "Studies of interference in serial verbal reactions," *Journal of Experimental Psychology*, vol. 18, 1935.

[9] F. H. Durgin, "The reverse Stroop effect," *Psychonomic Bulletin and Review*, vol. 7, 2000.

[10] H. Terakawa, H. Kato, Y. Yonemura, and Y. Katori, "Analysis of predictive coding model with hierarchical reservoir computing for modeling Stroop effects," *Proceedings - International SoC Design Conference 2023, ISOCC 2023*, 2023.

[11] H. Lim, Y. Wang, Y. Xiao, M. Hu, and D. J. Felleman, "Organization of hue selectivity in macaque V2 thin stripes," *Journal of Neurophysiology*, vol. 102, 2009.

[12] T. Smith and J. Guild, "The C.I.E. colorimetric standards and their use," *Transactions of the Optical Society*, vol. 33, 1931.

[13] D. Gabor, "Theory of communication," *Journal of the Institution of Electrical Engineers - Part I: General*, vol. 94, 1947.

[14] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, 2008.

[15] M. Ichikawa, "Temporal illusions in visual perception," *Kougaku (in Japanese)*, vol. 39, no. 2, pp. 82–88, 2010.