Feature Discovery in Object Individuation

Shohei Hidaka
Kyoto University and JSPS Research Fellow
Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan
hidaka@cog.ist.i.kyoto-u.ac.jp

Jun Saiki
Kyoto University
YoshidaNihonmatsu-cho, Sakyo-ku, Kyoto 606-8501, Japan
saiki@cv.jinkan.kyoto-u.ac.jp

***Abstract*— Many developmental studies have pointed out the relationship between children's ability of object cognition and word learning. In this study, the relationship between object segregation and feature bias to objects in word generalization was investigated with a connectionist model. In particular, the model focused on building feature and object representation from simple visual images. Some previously proposed connectionist models could account for children's systematic feature bias in word generalization based on the frequency of word learning. However, the previous model can not clearly account for a special case that late talkers, who have smaller vocabulary than average, do not show typical bias. We suggested that the proposed model could account some key issues in object segregation and word generalization including the late talkers' pattern.**

***Index Terms*— Statistical learning, Object segregation, Shape bias, Late talkers.**

I. Introduction

Many developmental studies over the years have pointed out the relationship between forming object concepts and word acquisition [3]. This study considers the process of forming object representations, and proposes a model associating empirical facts about object segregation and word generalization to novel objects. We briefly review two previously proposed hypotheses involving the relationship between object concept and its categorization. After reviewing experimental facts of object segregation and novel word generalization, we introduce the mainly focused hypothesis, statistical learning account [16], for novel word generalization.

Luria and Yudovich [7] claimed that naming objects helped children to categorize them. According to them, naming an object makes children attend to the feature of the objects relevant to the particular category. Meanwhile, according to Piaget and Inhelder [10], children in the end of sensorimotor period (18 to 24 months old) begin to symbolize the real world, and can associate object symbols to linguistic symbols. Therefore, Luria and Yudovich [7] claimed that linguistic categorization leads to object concept formation, while Piaget and Inhelder [10] clamed that object concept formation leads to language acquisition. These two hypotheses do not have to be mutually exclusive, in fact, some studies supported both effects [12].

A. Fast mapping

The sensorimotor period is the same time when children's vocabulary increases rapidly, called "vocabulary spurt". From this period, children efficiently learn novel words even in a single presentation, and this non-trial-and-error learning is called fast mapping [4], [6], [17]. A well known example of fast mapping is shape bias, that children systematically generalize words for solid objects based on shape similarity [6]. This systematic bias in generalization is found in various entities such as objects, substances, animals, and syntax, thus, these prior knowledge (bias) to perceptual objects and linguistic structures are supposed to constrain the range of generalization of novel words. The bias in word generalization have been investigated by novel word generalization to feature manipulated objects, called novel word generalization task. For example, children prefer to generalize solid objects based on shape similarity and nonsolid substances based on material similarity [1], [17].

Recently, various developmental studies have shown empirical evidences involving the mechanism of novel word generalization to objects. Smith et al. [16] hypothesized the mechanism as statistical learning of attention to object features. According to this hypothesis, children learn associational relationships among labeling, linguistic structures, object properties. Children systematically attend to particular features of objects with particular cues, because they learn the association between the cues and the features. For example, children generalized novel nouns to solid objects based on shape similarity [1], [4], [17], because most solid objects have a particular rigid shape, are named based on shape similarity, and belong to count nouns linguistically [14]. Furthermore, connectionist models were proposed to account for the mechanism why shape bias emerges in word learning [1], [15]. In the models, the emergence of shape bias is mainly due to the difference of learning frequency of nouns (i.e. more count nouns are learned than mass nouns). Specifically, when objects are defined by shape and texture, bias in naming (learning of category A and B in Figure 1c works for attention to shape and disregard of texture) influences learner's generalization (i.e. whether shape- or texture- based) of a novel exemplar (represented by "novel" in Figure 1).

However, some empirical facts, that the models based on leaning frequency can not account for, remain unclear. For example, in long-term training with artificial categories, learning with shape-similar categories enhanced children's shape bias, meanwhile, learning with material-similar categories did not enhance children's material bias (i.e. no significant difference to no-learning control group) [15].

If the only factor was frequency of learning to specific association, not only shape but also material bias would be acquired after the training. The further empirical evidence is that late talkers, who are normal but have smaller vocabularies than age-mates, showed texture bias, even though control group showed shape bias [5]. To explain this with Figure 1c, late talkers learn smaller vocabulary than age-mate (e.g. learning only category A) and do NOT learn texture-based nouns more frequently than shape-based nouns (i.e. the late talkers' proportion of learned count and mass nouns was similar with control's [5]). However, the late talkers showed texture bias instead of weak or no shape bias.

A solution for these conflicting facts is to consider that shape is somehow special regardless of its learning frequency. In fact, young infants (4.5- to 11.5- month-olds), who do not produce words, tend to segregate objects based on their shape. In this study, we extended the statistical learning account to cover object segregation, and showed the model could account for the conflicting problem. The statistical learning account proposed in previous studies have focused on the relationship between linguistic naming and word generalization, as clamed by Luria and Yudovich [7]. In other words, the previous model assumed that objects and their features (shape, color, and texture) were independent symbolic representations (Figure 1c). Shape bias depends only on what feature were used for noun categorization (i.e. shape bias emerges because most of nouns were categorized by shape similarity). In this study, we investigated the process that infants symbolize "independent" objects and features from unsegmented context. In other words, as Piaget and Inhelder [10] pointed out, we investigated what feature bias emerges when representations of "objects" are built in segregation process (Figure 1d). Next we briefly review some key experimental facts of object segregation.

B. Object segregation in infancy

It makes sense that object individuation, to know what is an object, is necessary to categorize objects. On the other hand, strong relationship between individuation and categorization, that object categorization helps object individuation, was suggested. Furthermore, some studies suggested that young infants could segment chunks from ambiguous context based on the statistical structure. Needham et al. [9] investigated how prior experience to a segregated object influences infants' segregation of the other similar object presented in next scene. 4.5-month-olds did not generalize an experience of a segmented object to segregation of the other objects with different texture [8], but they generalized an experience of *three* segregated objects with different texture from each others to segregation of the other objects with different texture [9]. Needham et al. [9] showed, with well manipulated control experiments, that young infants form object categories consisting of three objects, and the categorization influence object segregation. Furthermore, the feature usage for object segregation develops in order, shape and size (4.5 month), texture (7.5 month), color (11.5 month) [18]. Therefore, these infant's object segregation studies seem to support our conjecture, that shape is somehow special regardless of word learning, and that categorization and segregation of objects interact each other.

How do infants segregate objects? Empirical evidences were reported that infants could learn chunks embedded in context based on statistical information. Some developmental studies suggested that infants could extract chunks from auditory sequences or visual scenes without explicit labeling. For example, 8-month-old infants found artificial words from auditory sequences based on transitive probability [13], and 9-month-old infants found a frequently presented pair of objects from visual scene with some other pairs based on conditional probability [2]. In summary, word acquisition and object segregation shares statistical learning as their junction. In the next section, we propose a model building feature structure for object segregation, which covers key issues reviewed above.

II. Simulation

The purpose of the simulation is to investigate the relationship between object segmentation process and bias to feature defining the object. Therefore, we ignored or simplified both discriminability of realistic feature and learning frequency of objects and features, and focused on integration of perceptual features of objects in a simple visual model. More specifically, a visual scene with object defined by "shape", "color", and "texture" was presented (Figure 1a). We simulated association of the visual images to object representation defined by symbolic features such as "a bright circle with check pattern". Even though a simple visual object set was used, the object set simulated essential points identical to realistic structure, latent feature hierarchy. The feature hierarchy argued here is that shape is defined by relation of texture or color to background and texture is defined by relation of color (Figure 1b).

The statistical learning model [1], [15] simplified that object features (e.g. shape, color, texture, and solidity) were independent, and it focused on different frequency of categorization types (i.e. learning with category A and B causes shape bias in Figure 1c). However, the simulation here focus on prior process, the feature symbolization regardless of learning frequency, in order to extend the model to cover not

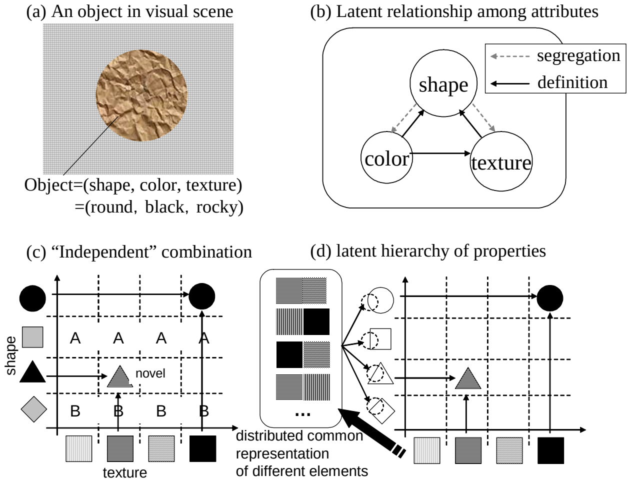

Fig. 1. Feature representation of object: (a) an object in visual scene parsed to shape, color and texture, (b) shape defined by color or texture contrast segregates what color and texture is, (c) internal object representation as combination of "independent features", (d) independent-appeared features share common representation.

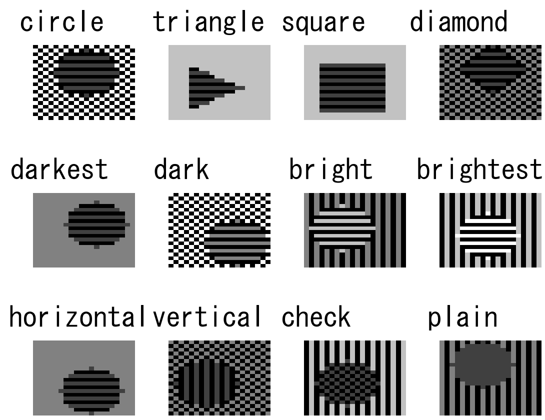

Fig. 2. Visual objects (retinal images) defined by shape, color, texture

only object-name association but also object-feature association. Therefore, we mainly investigated how an associator builds inter-feature relationship in object segregation when the objects are categorized uniformly (i.e. all cells in Figure 1d are learned equally). Different from the model with independent features (Figure 1c), in the object segregation, some features can share a common representation (left box in Figure 1d). In particular, the shared common representation could be more efficient coding objects when shape consists of relation among elemental features, such as texture or color (Figure 1b). In this study, we argue that this object building process is essential in both object segregation and object categorization in early developmental stage. Thus the proposed model investigated whether it could account for qualitative pattern of the late talker's bias in novel word generalization [5] and feature usage in object individuation [18]. The task in simulation was to output symbolic feature category from an object embedded in visual scene (Figure 3). We focused only on how much objects could be associated to symbolic feature in its segregation process, and ignored the other factors such as difference of learning frequency of features or categories, which was considered in previous study [1], [15]. In the simulation, well discriminated (associated) feature was considered as the feature which was sensitive in object individuation and object categorization. Notice that we do not argue pure feature discrimination but building of conjunct feature representation (i.e. "object").

A. Procedure

Visual objects in retinal field were defined by image with 20 by 20 pixels (Figure 2). The objects had different shape, color and texture pattern from the background. The task was to discriminate an object's subclasses of shape, color, and texture. The location of the visual objects was randomly decided not to have them out of the visual field. Specifically, a pixel consisted of a value (color) out of five (0, .25, .5, .75, 1), and the color pattern defines texture subclass (vertical stripes, horizontal stripes, check, plain), and the contrast of texture or color to background defines shape subclass (circle, triangle, square, diamond). The object set consisted of 1728 patterns (i.e. possible combination of shape (4), color (4), texture (4), background color (3), background texture (3) , random center location (3)). The visual images were filtered by gabor functions defined by equation below as early visual process.

$$G(x) = \exp(-\frac{x^2}{\sigma^2}) \cos(2\pi f(x-\theta)) \qquad (1)$$

where f and θ are frequency and phase, and $x = \sqrt{(y-y_0)^2 + (z-z_0)^2}$ is distance from the center (y_0, z_0) to a particular (y, z). The set of gabor functions consisted of 216 filters (orientation (4: $0°, 45°, 90°, 135°$), frequency (3: $0.5, 1.0, 2.0$) , size of receptive field (2), location of receptive field (9)). Through these filters, approximate 15.3% information of the original image set was lost (15.7% in no-background condition) [1]. After filtering the images by gabor functions, the signals were compressed by Principal Component Analysis (PCA) for ignoring 5% noise information , and we obtained input vectors with 47 and 51 dimension in no-background condition and the other condition (see

[1]The lost information was estimated by the coefficient of determination of K to I, where $J = GI$ and $K = (G^T G)^{-1} G^T J$. I and K are 400-dimension column vectors representing an original images and the image inversed from gabor signal J (216-dimension column vector) with G (216-by-400 matrix representing gabor filters).

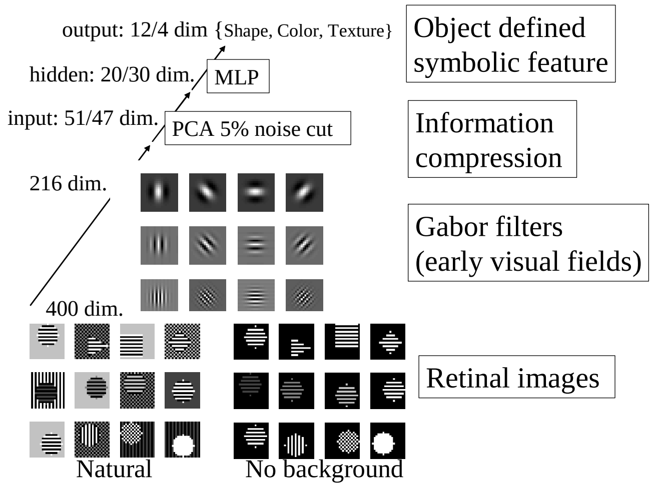

Fig. 3. Schematic image of simulation

also the next section), respectively. Therefore, total information taken away from the original images after PCA were approximately 20%. The PCA process means cognitively hebbian learning, unsupervised association, and technically a preprocess just for smooth learning of Multi-Layer Perceptron (MLP). The signals obtained from PCA were used as input for MLP , and the inputs were associated to symbolic feature, shape (e.g. circle) , color (e.g. darkest) and texture (e.g. horizontal stripe). Specifically, the output was local representation (e.g. (shape, color, texture)= (circle, bright, horizontal stripe)=(1,0,0,0,1,0,0,0,1,0,0,0)). Correct ratio of four subclasses in three features (i.e. discriminability) in each learning epoch was measured as bias towards each feature. Correctness was defined by match in each feature (four subclasses) between the one values in a teacher vector and the maximum value out of output vectors. This learning design is also interpreted that 64 objects (represented by distributed three features with four subclasses) are categorized (named) with equal frequency (i.e. all cells in Figure 1d are uniformly learned), against biased categorization (Figure 1c). Therefore, the correct ratio of subclass in three features is independent of its learning frequency of categories (i.e. main factor of shape bias in previous model [1], [15]), but it depends only on how the model built object representation. The chance level of the correct ratio in each feature is 25%. Thus some networks with lower correct ratio than 30% were considered as training failure and were excluded in analysis. The criterion was chosen, because the probability of the correct ratio higher than .3, when the model responds randomly out of four options 1728 times, is smaller than 10^{-3}.

B. Condition

There were 4 conditions, “base condition”, “single-feature condition”, “large-capacity condition”, and “no-background condition”. The base condition was the basis of all simulations, and the simulation in the other conditions had a particular different parameter (Table I). In the base condition, MLP with three layers (51 input units, 20 hidden units, and 12 output units) was trained to a set of objects 400 times by batch processing (i.e. one training to all patterns is called one epoch, thus the network was trained 400 epochs). Simulation in all conditions consisted of 100 networks with different initial states generated randomly.

The single-feature condition, the output of MLP was changed to single feature (i.e. only four subclasses of shape, color, or texture, thus there were four output units) on purpose to investigate baseline discriminability of each feature in this particular set. If feature representations in base condition were independent to each other, the performance of base condition and single-feature condition would be the same or proportional.

In no-background condition, the input of MLP was changed to images with single kind of background consisting of a dark color and plain texture (“No background” in Figure 3). The purpose of this condition was to investigate how object segregation influences on feature bias pattern. In this condition, the model does not have to segregate object in order to discriminate color and texture, because the color and texture of background was fixed in any training exemplars. Thus, in other words, the object segregation process in this condition was broken (the broken arrow in Figure 1b).

In large-capacity condition, the number of hidden units of MLP was changed to 30. The hidden units in MLP are supposed to represent relationship between image and symbolic feature, thus, the MLP with more hidden units has less loading on memory capacity. The goal of this condition is to investigate how the memory capacity for object segmentation influence feature bias.

condition	input	hidden	output	No. of successes
base	51	20	12	71
single-feature	51	20	4	73
no-background	47	20	12	68
large-capacity	51	30	12	73

TABLE I

THE NUMBER OF UNITS OF MLP AND SUCCESSFUL NETWORKS

III. RESULTS AND DISCUSSION

Two-factor analysis of variance (ANOVA) (37 levels of learning epochs (10 epoch interval from 40 to 400 epoch, except for the first 30 epochs with mean correct lower than .3) by three features) to correct ratios in base condition revealed the significant main effect of learning epochs (F(36, 10989)=22.6, $p < 10^{-3}$) and features (F(2, 10989)=545.8, $p < 10^{-3}$) but no significant interaction (F(72, 10989)=1.04, p=.387). This result indicated that the networks were trained successfully, that the correct ratios of three features were different, but that the relative correct patterns of three features did not vary along the learning epoch (after the first 30 learning epochs). Thus it also suggests that a network with higher performance in shape met a particular criterion earlier (Figure 5). Therefore, we identified correct ratio in the final epoch as developmental order of feature usage in object segregation. The correct ratios of simulation in each condition are shown in Figure 4. The error bars indicate the half standard deviation, and the asterisks indicate significant difference ($p < .05$) in one-factor ANOVA to correct ratio of each condition. The Tukey's multiple comparison test to the result in base condition showed significant differences in all three pairs of performances (i.e. combinations of shape, color and texture). This result indicates that the mean order pattern of feature bias was consistent to the order of children's feature usage for object segregation [18] (shape at 4.5 month, texture at 7.5 month, and color at 11.5 month).

Moreover, two-factor ANOVA (100 networks by three features) to correct ratios at 30 to 400 epoch in base condition revealed the significant main effect of individual difference (F(99, 10800)=256.8, $p < 10^{-3}$), features (F(2, 10800)=8.15, $p < 10^{-3}$), and significant interaction (F(198, 10800)=226.0, $p < 10^{-3}$). The result indicated that the correct pattern depended on individual difference of networks, thus we analyzed the pattern of individual networks sorted by the mean correct ratio of three features. The sorted performances were smoothed by mean with the closest 10 networks (e.g. the 50th-higher performance was substituted with mean of 45th- to 55th- higher performance) (left sides of Figure 6). The analysis showed the different structure of the networks in base and large-capacity condition. Most of the networks in base condition consistently had higher shape performance than texture (right upper graph in Figure 6) , and texture than color, meanwhile, the worse group of the networks in large-capacity condition had reversely higher texture performance than shape (right bottom graph in Figure 6). The detail analysis, shown in right sides of Figure 6, revealed that the performances subtracted texture from shape were positive in any networks in base condition, and those were negative and positive in a group with lower and higher performance in large-capacity condition. The correlation between mean performances and the subtracted values in base and large-capacity condition were respectively -.226 ($p = .055$) and .692 (significantly different from zero, $p < .001$). This result indicates that the networks basically had shape dominant performance, but that some poor learners with large capacity had texture dominant performance. This is consistent with texture bias found in late talkers [5]. Jones [5] showed that late talkers, normal children with relatively small vocabularies, tended to generalize novel words to objects based on texture similarity.

Generally speaking, the MLP with large number of hidden units definitely have better performance to training set than that with smaller number of hidden units. However, this is not simply better in terms of generalization to novel inputs, because some networks with large capacity are overfitted to a training set and tends to lose generalization to a novel set. Thus, the poor learners with large capacity in the model might reflect that late talkers tend to memorize peripheral feature of known objects and have trouble in its generalization to unknown ones. In fact, a study about the relationship between memory capacity and categorization suggested that infants with high memory load can grasp only central feature and infants with smaller memory load grasp both central and peripheral feature [11]. The different bias in base and large capacity condition might reflect this difference found in the developmental study. The networks with smaller capacity (i.e. high memory load) develop shape as central representation of objects , meanwhile the networks with large capacity (i.e. smaller memory load) were occasionally trapped by building peripheral representation (i.e. texture).

In single-feature condition, when the network associated objects to only single feature, the performance of each the three single features showed no significant difference (F(2, 210)=.83, p=.44). This result indicates baseline difficulty of each feature discrimination was not different. Therefore, we could reject the possibility, that the result in base condition reflected just difference of feature in a specific set. In stronger interpretation, this result might suggest that making representation of three features in conjunct form (like base condition) constrained to have shape bias, because representation for only single feature (like single-feature condition) did not show significant shape bias. In other words, the simulation reflected the association of dependent feature (unlike independent features in Figure 1c) to object, which was described in Figure 1d. In no-background condition, when the network associated all input images with the same background, showed significant difference in one-way ANOVA to correct ratios (F(2, 216)=6.0, p <.005). Further analysis with Tukey's multiple comparison test showed significant difference in the pairs shape/color and texture/color. The model in this condition did not have to "segregate" object from background because of constant background. Therefore, the comparison of performance in this condition to that in base condition suggested that object segregation would be critical whether shape bias emerged. In other words, the pressure, "segregating object", makes a learner attend to not texture but shape as representation of "object". In that sense, the object in no-background condition seems like a nonsolid substance which does not have rigid shape. Thus, these results might reflect that children tend to generalize novel word given to nonsolid substance base on texture or material similarity [1], [17].

In summary, the proposed model, which extended statistical learning model to object-feature association, reproduce

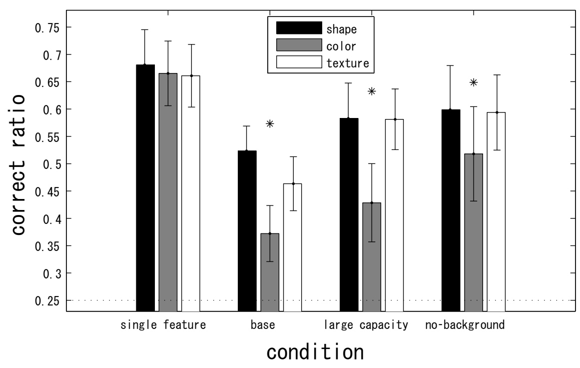

Fig. 4. Correct ratio

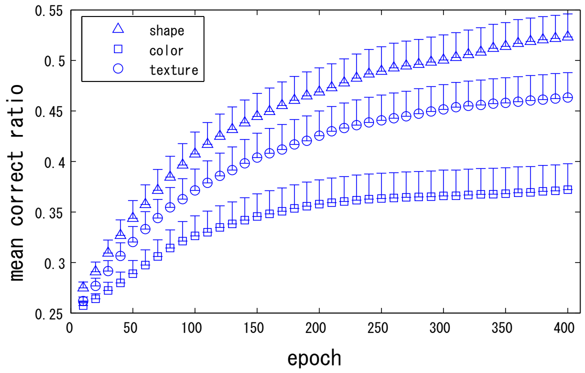

Fig. 5. Correct ratio in base condition along learning epoch

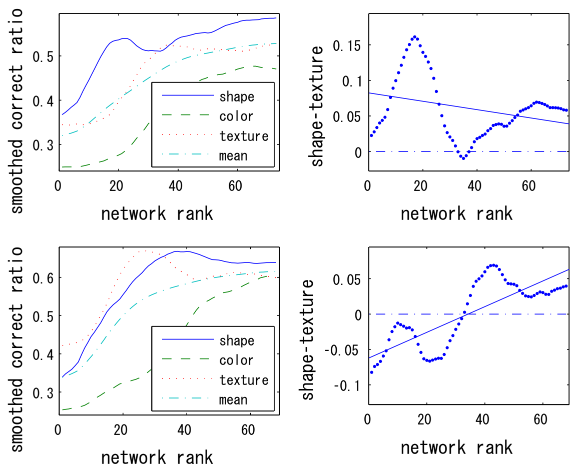

Fig. 6. Corret ratio (left) and the difference of its shape and texture (right) in base (upper) and large-capacity (bottom) condition smoothed and sorted by the mean performance.

qualitative pattern (in base condition) of shape bias found in novel word generalization [6] and development order of feature usage for object segregation [18]. Moreover, the model with large capacity also showed late talkers' texture bias [5] as individual difference of a group of networks (i.e. "good" or "poor" learners) regardless of learning frequency. According to the comparison of simulations in the four conditions, the emergence of shape bias and late talker's pattern in this model depends on capacity of internal representation, feature conjunction, and object segregation.

ACKNOWLEDGMENTS

This work was supported by Grants-in-Aid for Scientific Research from JMEXT (No. 15650046), JSPS Research Fellowships for Young Scientists, and the 21st Century COE Program from JMEXT (D-2 to Kyoto University).

REFERENCES

[1] E. Colunga and L. B. Smith, "From the Lexicon to Expectations About Kinds: A Role for Associative Learning", Psychological Review, vol. 112, no. 2, pp. 347-382, 2005.
[2] J. Fiser, and R. N. Aslin, "Statistical learning of new visual feature combinations by infants", Proceedings of the National Academy of Sciences, vol. 99, no. 24, pp. 15822-15826, 2002.
[3] A. Gopnik and A. Meltzoff, "Categorization and Naming: Basic-Level Sorting in Eighteen-Month-Olds and Its Relation to Language", Child Development, vol. 63, p.p. 1091-1103, 1992.
[4] M. Imai, and D. Gentner, "A cross-linguistic study of early word meaning: universal ontology and linguistic influence", Cognition, vol. 62, pp. 169-200, 1997.
[5] S. S. Jones, "Late talkers show no shape bias in object naming", Developmental Science, vol. 6, no. 5, pp. 477-483, 2003.
[6] B. Landau, L. B. Smith and S. S. Jones, "The Importance of Shape in Early Lexical Learning", Cognitive Development, vol. 3, pp. 299-321, 1988.
[7] A. R. Luria and F. Yudovoch, "Speech and development of mental processes in the child", London: Staples, 1959.
[8] A. Needham, "Object Recognition and Object Segregation in 4.5-Month-Old Infants", Journal of Experimental Child Psychology, vol. 78, pp. 3-24, 2001.
[9] A. Needham, G. Dueker and G. Lockhead, "Infants' formation and use of categories to segregate objects", Cognition, vol. 98, pp. 215-240, 2005.
[10] J. Piaget and B. Inhelder, "The psychology of the child", New York: Basic, 1969.
[11] P. C. Quinn, "The categorical representation of visual pattern information by young infants", Cognition, vol. 27, pp. 145-179, 1987.
[12] G. C. Roberts and K. N. Black, "The effect of naming and object performance on toy preferences", Child Development, vol. 43, p.p. 858-868, 1972.
[13] J. R. Saffran , R. N. Aslin and E. L. Newport, "Statistical learning by 8-month-old infants", Science, vol. 274, p.p. 1926-1928, 1996.
[14] L. Samuelson and L. Smith, "Early noun vocabularies: do ontology, category structure and syntax correspond?", Cognition, vol. 73, pp. 1-33, 1999.
[15] L. Samuelson, "Statistical Regularities in Vocabulary Guide Language Acquisition in Connectionist Models and 15-20 Month Olds", Developmental Psychology, vol. 38, pp. 1016-1037, 2002.
[16] L. B. Smith, S. S. Jones, B. Landau, L. Gershkoff-Stowe and L. Samuelson, "Object name learning provides on-the-job training for attention", Psychological Science, vol. 13, p.p 13-19, 2002.
[17] N. N. Soja, S. Carey, and E. S. Spelke, "Ontological categories guide young children's inductions of word meanings: object terms and substance terms", Cognition, vol. 38, pp. 179-211, 1991.
[18] T. Wilcox, "Object individuation: infants' use of shape, size, pattern and color", Cognition, vol. 72, pp. 125-166, 1999.