Cañamero, L., Oudeyer, P.-Y., and Balkenius, C. (Eds) Proceedings of the
Ninth International Conference on Epigenetic Robotics: Modeling Cognitive
Development in Robotic Systems. Lund University Cognitive Studies, 146.

Caregiver's Auto-mirroring and Infant's Articulatory Development Enable Vowel Sharing

Hisashi Ishihara* **Yuichiro Yoshikawa**** **Minoru Asada****,*
*Graduate School of Eng., Osaka University
**Asada Synergistic Intelligence Project, ERATO, JST
2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan
hisashi.ishihara@ams.eng.osaka-u.ac.jp, yoshikawa@jeap.org, asada@ams.eng.osaka-u.ac.jp

Abstract

We extend the auto-mirroring guidance model, which explains the process of sharing vowels between a caregiver and an infant, by introducing two transitional elements related to the infant's articulatory development: One is the accuracy of the infant's articulation improving along with the separation of his/her vowel prototypes. The other is the transition of the caregiver's auditory perception of mapping the infant's vowels onto her own ones. The extended model can simulate several additional aspects of vowel development, e.g., the rapid separation of infant vowels and their convergence and the transient rise of stretching motherese. Simulation results suggest a new picture of the process of vowel development, which explains how there are two transitional aspects of vowel separation and guidance, and they also suggest hypotheses on the causes of vowel separation and a caregiver's motherese.

1. Introduction

The process of sharing vowels with caregivers seems to be the first developmental step of an infant's language development. Kuhl and her colleagues pointed out the importance of regarding the process of vowel development as dynamic interaction between perceptual development, articulatory development, and a caregiver's address to the infant(Kuhl et al., 2008). However, their model is still too conceptual to understand the computational mechanism underlying such processes of development. To reveal such a mechanism, the use of synthetic studies has been considered one of the most promising approaches (Asada et al., 2009).

Some of these studies have focused on the perceptual development needed to learn a caregiver's vowel categories from her speech (McMurray et al., 2009, Vallabha et al., 2007). However, how the caregiver addresses her infant in speech should be taken into account to understand such perceptual development. This addressing by the caretaker seems to depend on her observations and understanding of the infant's developmental stage, based on such input as the quality of vocalizations (Gros-Louis et al., 2006, Bloom and Lo, 1990).

Mutual imitation (Masur and Olson, 2008, Kokkinaki and Kugiumutzakis, 2000) is a typical and highly significant instance of caregiver-infant interaction. de Boer (de Boer, 2000) and Oudeyer (Oudeyer, 2005) have suggested that imitative interaction plays important roles in sharing vowels between agents. In particular, Oudeyer showed computationally that shared prototypes can be self-organized by virtue of a perceptual bias around vowel prototypes (perceptual magnet effect (Kuhl, 1991)). However, this research lacks consideration of an inevitable hurdle to infant development, namely the physical differences between the caregiver and infant: They cannot produce the same vowel sounds since their articulatory organs are very different from each other (Vorperian and Kent, 2007).

Miura *et al.* showed that a robot could acquire shared vowels with a human interactant who imitates its vocalization with a different articulatory organ from it (Miura et al., 2007). To investigate what properties of a caregiver's imitation permit the sharing of vowels, Ishihara *et al.* have constructed a computational model of the caregiver-infant imitative interaction (Ishihara et al., 2008). They proposed that an infant's prototypes are guided toward the anticipated ones by distinctive biases in the caregiver's imitation, made as if she were imitating not only the infant's utterance but also both her own usual utterance style (sensorimotor magnet bias) and her own previous utterance (auto-mirroring bias). Also, they considered the differences in utterable vowels. However, there remain other aspects of infant articulatory development such as proficiency of articulation control through self-monitoring experience of produced sound (Oller and Eilers, 1988) and expansion of one's utterable vowel area in vowel space (Ishizuka et al., 2007, Rvachew et al., 2006). Such immature articulation activities should be introduced in their model, given the likelihood that the developmental conditions of perception and articulation could affect each other (Vihman and Nakai, 2003, van Beinum et al., 2001, Oller and Eilers, 1988).

Introducing the elements of an infant's articulatory development in the auto-mirroring guidance model would allow us to examine the causes and effects of several phenomena appearing in a real infant's vowel development, such as 1) rapid separation of distribution clusters of infant utterances and its subsequent convergence (Ishizuka et al., 2007) and 2) stretching motherese in which the distributional profile of the mother's vowels addressed to her infant tends to stretch compared to that used in addressing other adults (Kuhl et al., 1997).

In this paper, we extend the auto-mirroring guidance model by introducing two transitional elements related to an infant's articulatory development. First, the accuracy of an infant's articulation is assumed to improve along with the separation of the infant's own prototypes, since sensorimotor learning of these prototypes would be easier after they are separated more widely. Then, the caregiver's auditory perception that maps the infant's vowels on the caregiver's own vowels is also modulated according to the infant's articulatory development. We report that the extended model can simulate the rapid separation of an infant's vowels and their convergence as well as the transient rise of stretching motherese. Furthermore, we suggest a new picture of the process of vowel development that explains that there are two transitional aspects, i.e., separation and the guidance, and we suggest hypotheses on the causes of infant vowel separation and the caregiver's use of motherese.

2. Auto-mirroring guidance model

2.1 Overview

This model consists of imitation mechanisms for both a caregiver and an infant and a learning mechanism of a sensorimotor map for the infant. Imitation mechanisms convert the other's vowel sound into the imitator's own articulation command to produce the imitation vowel sound.

Another feature of this mechanism is to contain possible biasing elements, i.e., sensorimotor magnets and auto-mirroring bias, in the caregiver's imitation arising from her anticipation of her infant's utterance. Sensorimotor magnets are kinds of convergence bias of perception and articulation around the caregiver's vowel prototypes. Part of this characteristic seems to originate from Kuhl's perceptual magnet effect (Kuhl, 1991), which is a perceptual warp around a listener's phoneme prototypes. This bias can be seen as an effect of the caregiver's unconscious anticipation of her infant to articulate vowel prototypes in a mother language. Another bias, auto-mirroring bias, is a kind of mixing bias of the other and the self, in which the perception of the other's vowel is attracted toward the perceiver's own last utterance. This bias can be seen as an effect of the caregiver's anticipation that her infant will imitate her utterance correctly.

Figure 1 shows an overview of the model. At the t-th step of interaction, the infant utters a vowel $\boldsymbol{s}'(t) \in \Re^{N_s}$ by the articulation command $\boldsymbol{a}'(t) \in \Re^{N_a}$, and the caregiver listens to vowel sound $\boldsymbol{s}'(t)$ and utters $\boldsymbol{s}(t) \in \Re^{N_s}$ by the articulation $\boldsymbol{a}(t) \in \Re^{N_a}$ as an imitation of $\boldsymbol{s}'(t)$. Next, the infant listens to $\boldsymbol{s}(t)$ and updates his/her sensorimotor map based on both his/her articulation $\boldsymbol{a}'(t)$ and the caregiver's reply $\boldsymbol{s}(t)$ and then tries to imitate $\boldsymbol{s}(t)$ by the articulation $\boldsymbol{a}'(t+1)$ using the updated map. The learning mechanism of the map is explained in section 2.3 below.

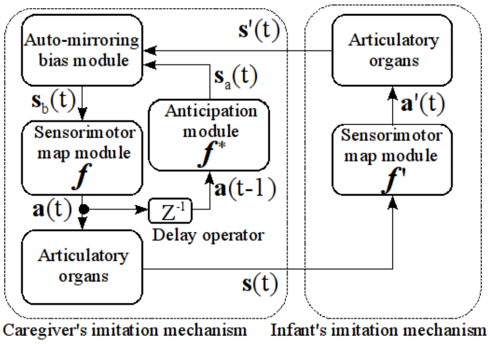


Figure 1: Overview of proposed model consisting of two imitation mechanisms for a caregiver and an infant.

2.2 Imitation mechanism

The caregiver's imitation $\boldsymbol{a}(t)$ is modeled by the formula:

$$\boldsymbol{a}(t) = f(\boldsymbol{s}_b(t); \boldsymbol{p}_i, \lambda), \tag{1}$$

$$\boldsymbol{s}_b(t) = (1-\eta)\boldsymbol{s}'(t) + \eta \boldsymbol{s}_a(t), \tag{2}$$

$$\boldsymbol{s}_a(t) = f^*(\boldsymbol{a}(t-1); \boldsymbol{p}_i^*), \tag{3}$$

where $f : \Re^{N_s} \rightarrow \Re^{N_a}$ is her sensorimotor map at the t-th step based on her prototypes $\boldsymbol{p}_i \in \Re^{N_a}$, while $f^* : \Re^{N_a} \rightarrow \Re^{N_s}$ is its quasi-inverse map at the t-th step based on the infant's anticipated prototypes $\boldsymbol{p}_i^* \in \Re^{N_s}$. This term is used because $\boldsymbol{p}_i^*$ represents the infant's producible vowels that can be mapped onto the caregiver's prototypes, i.e., the caregiver anticipates her infant matching her own prototypes $\boldsymbol{p}_i'$ with $\boldsymbol{p}_i^*$.

The caregiver's articulation $\boldsymbol{a}(t-1)$ is input to her quasi-inverse sensorimotor map f^* and converted to the anticipation $\boldsymbol{s}_a(t)$ of her infant's imitation of the articulation $\boldsymbol{a}(t-1)$. The anticipation is mixed with real infant utterance $\boldsymbol{s}'(t)$ with the mixing rate $\eta (0 \leq \eta \leq 1)$, and this attraction is called the auto-mirroring bias. Then the attracted perception $\boldsymbol{s}_b(t)$ is converted to the articulation command $\boldsymbol{a}(t)$ by the sensorimotor map f, where its output is attracted to prototypes $\boldsymbol{p}_i (i = 1, \cdots, M)$ with the converging degree $\lambda (0 \leq \lambda \leq 1)$. This convergence is called the sensorimotor magnets. Thus, we can control the strength of the caregiver's biases, that is, the sensorimotor magnets and auto-mirroring bias, by changing the parameters λ and η.

2.2.1 Sensorimotor map

We model the sensorimotor map f with one of the linear regression mixture models, a Normalized Gaussian Network (NGnet) (Sato and Ishii, 2000, Moody and Darken, 1989). An NGnet has M Gaussian functions $g_i (i = 1, \cdots, M)$ as basis functions in input space and maps the input data with mixed linear regression functions. Each mixture rate is decided according to the distance between the input and the center of each Gaussian, each of which has charge of one linear regression function. NGnet determines the caregiver's articulation $\boldsymbol{a}(t)$ by

$$\boldsymbol{a}(t) = f(\boldsymbol{s}_b(t); \boldsymbol{p}_i, \lambda) \tag{4}$$

$$= \sum_{i=1}^{M} \frac{g_i(\boldsymbol{s}_b(t); \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{j=1}^{M} g_j(\boldsymbol{s}_b(t); \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \boldsymbol{W}_i(\boldsymbol{p}_i, \lambda, \boldsymbol{\mu}_i) \boldsymbol{s}_b(t), \tag{5}$$

where $\boldsymbol{\mu}_i \in \Re^{N_s}$ and $\boldsymbol{\Sigma}_i \in \Re^{N_s \times N_s}$ are the center vector and the variance-covariance matrix of the i-th Gaussian. λ is a parameter that sets the eigenvalue of the representation matrix of linear transformation $\boldsymbol{W}_i(\boldsymbol{p}_i, \lambda, \boldsymbol{\mu}_i) \in \Re^{N_a \times (N_s+1)}$ to $(1-\lambda)$. Furthermore, $\boldsymbol{s}_b(t) \equiv [\boldsymbol{s}_b^{\mathrm{T}}(t), 1]^{\mathrm{T}} \in \Re^{N_s+1}$ is the argumented matrix of $\boldsymbol{s}_b(t)$.

Figure 2 shows how sensorimotor magnets are modeled and controlled by the setting of λ when we assume that the NGnet has one Gaussian unit $(M = 1)$, where the one-dimensional inputs (infant's vowel sounds) are normally distributed around its center. Each input is mapped by a matrix of linear transformation $\boldsymbol{W}_1$, and thus the distribution of the outputs (caregiver's articulation commands) are determined by the eigenvalue (slope here) of the matrix: The smaller the eigenvalue $(1-\lambda)$ of the transformation matrix $\boldsymbol{W}_1$ is, the more the distribution gathers around the image of the Gaussian center $\boldsymbol{\mu}_1$ under $\boldsymbol{W}_1$, namely $\boldsymbol{W}_1\boldsymbol{\mu}_1$. Therefore, we regard the image as a prototype to represent sensorimotor magnets, namely $\boldsymbol{p}_i \equiv \boldsymbol{W}_i\boldsymbol{\mu}_i$. Furthermore, we regard the center of Gaussian $\boldsymbol{\mu}_i$ as anticipated prototypes $\boldsymbol{p}_i^*$, namely $\boldsymbol{p}_i^* \equiv \boldsymbol{\mu}_i$, since they are mapped onto the caregiver's prototypes.

The caregiver's quasi-inverse sensorimotor map is also modeled by another NGnet so that it can map the caregiver's prototype $\boldsymbol{p}_i$ to $\boldsymbol{p}_i^*$ as opposed to the sensorimotor map that maps $\boldsymbol{p}_i^*$ to $\boldsymbol{p}_i$. This quasi-inverse map works like a predictor of the infant's imitation and is updated at every step as the sensorimotor map changes, as mentioned in section 3.2 below.

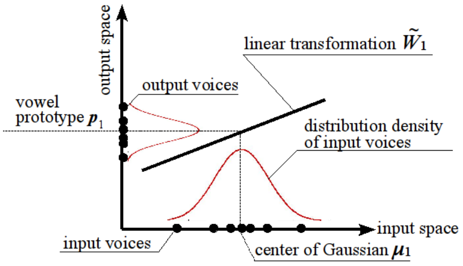


Figure 2: Illustration of how an NGnet constructs the sensorimotor magnets with one transformation matrix

2.3 Learning mechanism for an infant

An infant has the immature sensorimotor map f' represented by an NGnet that has M-Gaussian functions and learns its parameters in the T-th step of the interaction based on the n-step history $H(T)$ of the pairs of the infant's own articulation and the caregiver's reply. Namely, $H(T) = \{\boldsymbol{a}'(t), \boldsymbol{s}(t) | t = T - n + 1, \cdots, T\}$. Here, the infant task is to tune the parameters $\{\boldsymbol{\mu}_i'(T), \boldsymbol{\Sigma}_i'(T), \boldsymbol{W}_i'(T) | i = 1, \cdots, M\}$ of her sensorimotor map so that it can represent the input-output relationship from $\boldsymbol{s}(t)$ to $\boldsymbol{a}'(t)$ within $H(T)$. We use the EM algorithm (Sato and Ishii, 2000, Dempster et al., 1977), which is one of the maximum likelihood estimation methods for mixture models, to estimate the most appropriate parameters.

As a result of this update, the infant's prototypes $\boldsymbol{p}_i' \equiv \boldsymbol{W}_i'\boldsymbol{\mu}_i'$ are also updated at every step. The final goal of his/her development is to match his/her prototypes $\boldsymbol{p}_i'$ to his/her anticipated prototypes by the caregiver $\boldsymbol{p}_i^*$, which he/she can not observe directly.

3. Extensions of the model

3.1 Accuracy of infant's articulation

To consider the possible effects of the articulatory development of an infant on the process of sharing vowels, we introduce a simplified model. The accuracy of the infant's articulation control is considered to be improved through self-monitoring of the sound resulting from his/her attempts at articulation (Oller and Eilers, 1988). Therefore, this accuracy is modeled so as to be related to the current distribution of the infant's produced sounds: We assume that his/her articulation error can be represented by a variance of a Gaussian distribution around his/her target articulation $\boldsymbol{a}'(t)$, and this variance is determined based on the extent to which his/her current prototypes are separated from each other.

Given the infant's target articulation $\boldsymbol{a}'(t)$, the produced articulation is determined by

$$\boldsymbol{a}'(t) = \mathcal{N}(\boldsymbol{a}'(t), \sigma^2(S(t), h)), \tag{6}$$

$$\sigma(S(t), h) = \frac{150}{1 + \exp\{0.02(S(t) - h)\}}, \tag{7}$$

$$S(t) = \sum_{i=1}^{M} \Big(\frac{|\boldsymbol{p}_i'(t) - \sum_{j=1}^{M} \frac{\boldsymbol{p}_j'(t)}{M}|}{M} \Big), \tag{8}$$

where $\mathcal{N}(\boldsymbol{a}', \sigma^2)$ represents a manipulation to add a Gaussian noise whose variance is σ^2 to the target articulation $\boldsymbol{a}'$. The variance σ^2 depends on the degree of the prototypes' separation $S(t)$, and their relationship is controlled by the parameter h. Here, we can control the difficulty of articulation development in the simulation by changing the parameter h; for example, the larger h is, the larger σ^2 is, even under the same condition of $S(t)$.

3.2 Transitions of caregiver's perception

In the previous model, we assumed that anticipated prototypes $\boldsymbol{p}_i^*$ are fixed throughout interactions. However, an experimental result of category identification of infants' vowels by caregivers shows that an adult's perception of infant vowels can alter, e.g., the geometry of perceived prototypes expands and shifts in vowel space as the infant becomes older (Vorperian and Kent, 2007, Ishizuka et al., 2007, Rvachew et al., 2006, Kuhl and Meltzoff, 1996). Therefore, anticipated prototypes should be altered through interactions for a more valid simulation.

We introduce the expansion of the geometry of anticipated prototypes in a fixed expansion rate as a first implementation. Anticipated prototypes are determined by

$$\boldsymbol{p}_i^*(t) = \boldsymbol{p}_i^*(0) + \frac{t}{T_L}(\boldsymbol{p}_i^*(T_L) - \boldsymbol{p}_i^*(0)), \qquad (9)$$

$$\boldsymbol{p}_i^*(0) = \sum_{i=1}^{M}(\frac{\boldsymbol{p}_i^*(T_L)}{M}), \qquad (10)$$

where T_L is the number of total interaction steps and $\boldsymbol{p}_i^*(T_L)$ are fixed values in the current model.

4. Simulation of mutual imitation

4.1 Procedure

A caregiver imitates her infant's utterance at every step while her infant basically tries to imitate the caregiver's utterance at every step but sometimes utters randomly, i.e., the infant tries to utter one of the prototypes every step until the n-th step and continues to do so every fifth step even after the n-th step interaction. Furthermore, until the n-th step has passed, the infant does not update his/her sensorimotor map, since she can not utilize enough learning data. In this simulation, we set $n = 500$ and $T_L = 5000$.

4.2 General settings

We assume each vowel sound is represented by a two-dimensional vector, since vowel prototypes are known to be distinguishable at two frequency peaks, which are called first formant and second formant. Furthermore, for simplicity of simulation, we assume that an articulation command can be represented by the same vector as that of the vowel sound produced by the articulation, i.e., $\boldsymbol{s}(t) = \boldsymbol{a}(t)$ and $\boldsymbol{s}'(t) = \boldsymbol{a}'(t)$.

Figure 3 shows an overview of the settings of the caregiver's prototypes $\boldsymbol{p}_i$ (blue dots), the anticipated prototypes $\boldsymbol{p}_i^*(t)$ (red dots), and the infant's prototypes $\boldsymbol{p}'_i(t)$ (black dots) in the vowel/articulation feature space. Anticipated prototypes are set to be located at a distance from the caregiver's prototypes, since their articulation organs are different and thus a difference in their vowels can be expected. We set the number of prototypes M to 5, imagining a Japanese caregiver. The infant's initial prototypes are set more closely together, since a real infant's categories are not so widely separated from each other in the early period of development. The caregiver's prototypes are fixed throughout the interactions, while anticipated prototypes are gradually expanded as the interaction proceeds based on eq. (9).

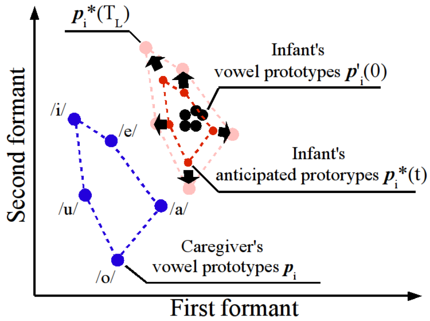


Figure 3: Overview of settings of the initial states of infant's prototypes (black dots) and caregiver's prototypes (blue dots) and the anticipated prototypes (red dots).

4.3 Settings of the caregiver

Assuming a Japanese caregiver and infant, we determined the caregiver's prototypes $\boldsymbol{p}_i$ and anticipated prototypes at the last step $\boldsymbol{p}_i^*(T_L)$ as

$$\{\boldsymbol{p}_i\} = \left\{ \begin{matrix} 700 \\ 1200 \end{matrix}, \begin{matrix} 400 \\ 1700 \end{matrix}, \begin{matrix} 400 \\ 1300 \end{matrix}, \begin{matrix} 600 \\ 1500 \end{matrix}, \begin{matrix} 500 \\ 1000 \end{matrix} \right\}, \qquad (11)$$

$$\boldsymbol{p}_i^*(T_L) = \boldsymbol{p}_i + \begin{matrix} 400 \\ 600 \end{matrix} \quad (i = 1, \cdots, M). \qquad (12)$$

We determined the parameters of the caregiver's sensorimotor map so that all of our assumptions are satisfied as follows:

$$\boldsymbol{\mu}_i(t) = \boldsymbol{p}_i^*(t) \quad (i = 1, \cdots, M), \qquad (13)$$

$$\boldsymbol{\Sigma}_i = \begin{matrix} 3600 & 0 \\ 0 & 3600 \end{matrix} \quad (i = 1, \cdots, M), \qquad (14)$$

$$\tilde{\boldsymbol{W}}_i(\boldsymbol{p}_i, \lambda, \boldsymbol{\mu}_i(t)) = \big((1-\lambda)\boldsymbol{I}, \boldsymbol{p}_i - (1-\lambda)\boldsymbol{\mu}_i(t)\big)$$
$$(i = 1, \cdots, M, 0.0 \leq \lambda \leq 1.0). \qquad (15)$$

In addition, we determined the parameters of the caregiver's quasi-inverse sensorimotor map so that all of our assumptions are satisfied as follows:

$$\boldsymbol{\mu}_i^* = \boldsymbol{p}_i \quad (i = 1, \cdots, M), \qquad (16)$$

$$\boldsymbol{\Sigma}_i^* = \begin{matrix} 3600 & 0 \\ 0 & 3600 \end{matrix} \quad (i = 1, \cdots, M), \qquad (17)$$

$$\tilde{\boldsymbol{W}}_i^*(\boldsymbol{p}_i^*(t), \boldsymbol{\mu}_i^*) = \Big(\boldsymbol{I}, \boldsymbol{p}_i^*(t) - \boldsymbol{\mu}_i^*\Big) \quad (i = 1, \cdots, M). \qquad (18)$$

Note that the infant's anticipated prototypes $\boldsymbol{p}_i^*(t)$ change through the interactions according to eq. (9).

From the simulation results of the previous model, we know that infant prototypes $\boldsymbol{p}_i'(t)$ are gradually guided toward the anticipated prototypes $\boldsymbol{p}_i^*(t)$ by virtue of association with the caregiver's sensorimotor magnets and auto-mirroring bias. The degree of such guidance depends on the degree of their strengths, and $(\eta = 0.5, \lambda = 0.6)$ is the setting pair that exerts the guidance effect most strongly. Therefore, we selected this setting pair for the current simulation.

5. Results

We simulated the caregiver-infant imitative interaction under several conditions of difficulty in articulatory development: (a)$h = 0$, (b)$h = 150$, (c)$h = 300$, and (d)$h = 450$.

5.1 *Fading of articulation error*

Figure 4 shows the degree of an infant's articulation error $\sigma(t)$ processed throughout interaction steps under each condition of h. We can see that the infant's articulation error $\sigma(t)$ is larger throughout the interactions under the condition where the infant's articulation development is more difficult (h is larger), as we had expected. In addition, we can see that the error $\sigma(t)$ tends to decrease step by step rapidly in the first half of this period, especially under conditions (a), (b), and (c).

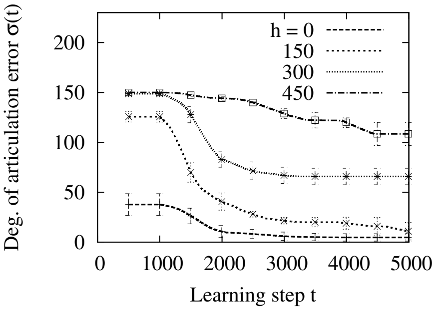
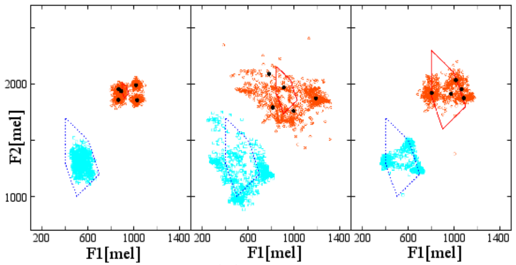


Figure 4: Differences in the transitions of infant articulation error $\sigma(t)$ under several conditions of difficulty h in articulation development.

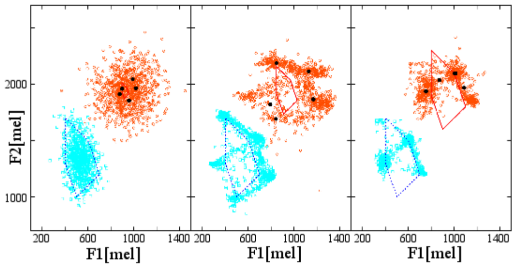
5.2 *Transitions of vowel distribution*

Figure 5 illustrates the transitions of vowel distributions of both the caregiver and the infant under each condition of h. In these figures, utterances of the infant $\boldsymbol{s}'(t)$ (red dots) and those of the caregiver $\boldsymbol{s}(t)$ (blue dots) and the infant's prototypes $\boldsymbol{p}_i'(t)$ (black dots) are plotted during each of three periods (at the last step for $\boldsymbol{p}_i'(t)$): first 1000 steps (left box), middle 1000 steps (middle box), and final 1000 steps (right box). The apexes of the red pentagons represent the infant's anticipated prototypes $\boldsymbol{p}_i^*(t)$ at the last step of each period, while those of the blue pentagons represent the caregiver's prototypes $\boldsymbol{p}_i$.

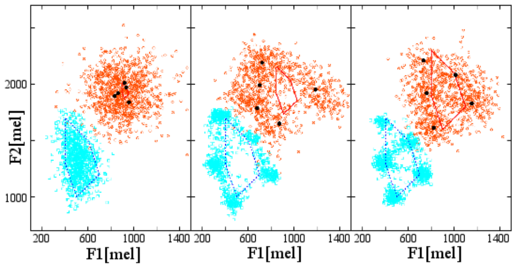
Large variations in the distributional patterns of utterances between conditions indicate that the difficulty of articulatory development heavily affected both the infant's learning and the interactions: The infant's utterances and prototypes were distributed more widely when h was larger and the number of uttered categories was different, i.e., only three vowel categories were uttered during the final period in condition (a) while five categories were uttered under conditions (c) and (d).



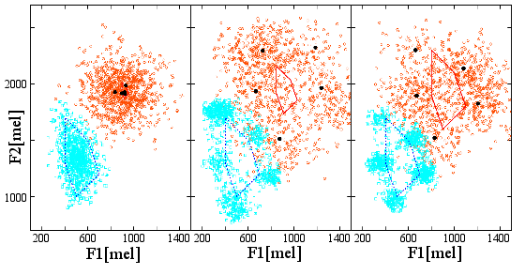
(a) $h = 0$



(b) $h = 150$



(c) $h = 300$



(d) $h = 450$

Figure 5: Transitions of vowel distributions of both the caregiver (blue dots) and the infant (red dots) and those of the infant's prototypes (black dots) under different conditions of h. The geometry of the caregiver's prototypes (apexes of blue pentagons) and the infant's anticipated prototypes (apexes of red pentagons) are also depicted.

5.3 Separation of prototypes

Ishizuka and Mugitani investigated the distributions of real infants' utterances during the age span of 4-60 months (Ishizuka et al., 2007). They showed that the geometry of vowel categories tend to expand until age 24 months and the speed of their separation is rapid in the early stage and then becomes slower.

Figure 6 shows the transitions of the separation degrees $S(t)$ of infant prototypes defined in eq. (8). The counterpart for the anticipated prototypes is also depicted as a reference by the solid line. We can see that the infant's prototypes tend to expand rapidly, particularly in the first half of the period, and then the speed of expansion becomes slower until the prototypes gradually converge. Interestingly, this basically reproduces the real transition reported in the previous study (Ishizuka et al., 2007).

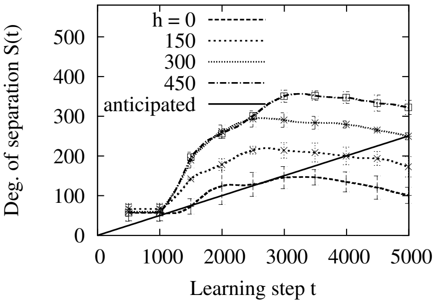


Figure 6: Transitions of the separation degree $S(t)$ of the infant prototypes under different conditions of h. The counterpart of the anticipated prototypes is also depicted.

5.4 Process of sharing vowels

Figure 7 shows the transition of the sharing degree of prototypes between the caregiver and the infant, which represents the as-a-whole closeness of the anticipated prototypes $\boldsymbol{p}_i^*(t)$ with the infant prototypes $\boldsymbol{p}'_i(t)$. The sharing degree is evaluated by the formula:

$$D(t) = \sum_{i=1}^{M} \frac{\mathrm{Min}(\{|\boldsymbol{p}_i^*(t) - \boldsymbol{p}'_j(t)|\}_{j=1,\cdots,M})}{M}, \qquad (19)$$

where $D(t)$ represents the as-a-whole distance of anticipated prototypes from the infant's prototypes at the t-th step of the interaction; consequently, the sharing degree is higher when this index is lower. The as-a-whole distance $D(t)$ is small in the initial state under all conditions, and these values are not so different from each other since the infant's prototypes are set to gather around the initial point of anticipated prototypes $\boldsymbol{p}_i^*(0)$. The as-a-whole distance $D(t)$ continued to increase under conditions (a) and (b) to the end, while it increased rapidly during the first half of the period and then began to decrease under conditions (c) and (d).

We can also see such transitions in Fig. 5. Under conditions (a) and (b), infant prototypes did not separate so widely from each other and therefore they were guided to a smaller number of anticipated prototypes, indicating that the large as-a-whole distance remained. On the contrary, under conditions (c) and (d), the geometry of infant prototypes expanded more widely than did that of anticipated prototypes in the first half of the period, and then the prototypes were located near all of the anticipated prototypes, indicating an inverted U-shape transition of the as-a-whole distance.

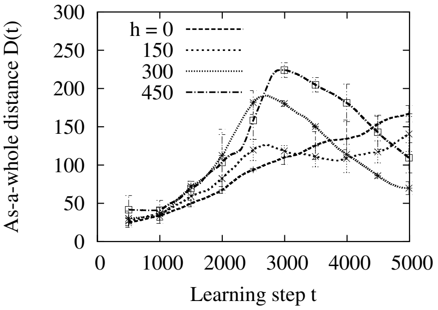


Figure 7: Transitions of the as-a-whole distance $D(t)$ of anticipated prototypes from infant prototypes under different conditions of h.

6. Discussion

6.1 Dominance of either of two aspects

There seems to be two aspects in the transitional process of sharing vowels: the separation of prototypes and the guidance of prototypes toward anticipated prototypes. The investigation in the previous work (Ishihara et al., 2008) revealed that the aspect of guidance can be caused by the balanced action of two biases in the caregiver's imitation, i.e., sensorimotor magnets and auto-mirroring bias. The other aspect, the separation of prototypes, can be considered a consequence of the infant's articulation error due to the larger error results in the larger distribution of the infant's utterances.

A simulated developmental process can be divided into two stages by focusing on which aspect is dominant. In the first stage, the aspect of separation is dominant due to the large articulation error. In this stage, prototypes become separated from each other because the separating effect surpasses the guiding effect at some point. The dominance of the aspect of separation becomes weakened gradually as prototypes become separated from each other, since an infant's articulation error decreases with the separation, as modeled in eq. (8). Then the next stage, where the aspect of guidance is dominant, begins. In this stage, the infant's prototypes can be guided toward anticipated prototypes as in the latter half of the period under conditions (c) and (d) shown in Fig. 7, since the guiding effect surpasses the separating effect this time at some point.

The prototypes can be effectively shared when the transition of these dominances is achieved appropriately: There seems to be adequate difficulty h of articulatory development, as under condition (c), where all prototypes can correspond to any of the anticipated prototypes. If the separating effect is too weak, as under conditions (a) and (b), the prototypes cannot sufficiently spread to correspond to all anticipated prototypes. On the other hand, if the separation effect is too strong, as under condition (d), the prototypes are not guided sufficiently, since the separating effect could cancel out the guiding effect.

6.2 Conditions for rise of motherese

We can find the rise of motherese particularly in the middle of the period in Fig. 5 (c). The left side of Fig. 8 illustrates how the geometry of centers of the caregiver's vowel clusters (blue pentagon with solid line) is stretched compared to that of her usual ones, or her prototypes (blue pentagon with dotted line). By comparing these results with those in the right side of Fig. 8, which shows the motherese reported by Kuhl *et al.* (Kuhl et al., 1997), we can see that this stretching property resembles the property of a real caregiver, i.e., the distribution of the caregiver's vowels addressed to infants (triangle with solid line) is stretched compared to that of vowels addressed to other adults (triangle with dotted line).

This characteristic seemed to arise in the simulation when the caregiver underestimated the degree of expansion of infant prototypes, in other words, when this degree exceeded that of the anticipated prototypes, as we can see in the left box of Fig. 8. In such cases, the caregiver perceives the infant's utterances as exaggerated ones and thus the caregiver's utterances also become exaggerated through her imitations.

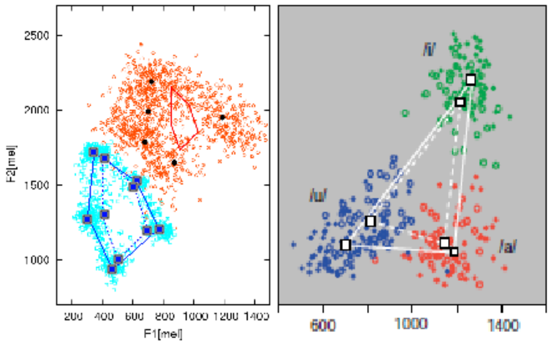


Figure 8: Stretching motherese in our simulation in the middle of the period under condition (c) (left box) compared to that reported by kuhl *et al.*, 1997 (right box).

7. Conclusion

According to the results obtained in this study, we suggest a more progressive picture of the developmental process of sharing vowels through vocal imitation as follows:

1. An infant's prototypes begin to separate from each other rapidly, since the separating effect caused by the infant's articulation error exceeds the guiding effect caused by the caregiver's biases. We can see this trend in the first half of the period in Figs. 5 and 6.
2. As the infant prototypes expand, he/she begins to utter vowels perceived by the caregiver as prototypes, and thus they come to utter more prototypical vowels as phonemes of their mother language. This trend can be seen in the first and middle parts of the period in Fig. 5.
3. The geometry of anticipated prototypes expands as the infant develops. Stretching motherese arises when the degree of expansion of the infant's prototypes exceeds the degree of the caregiver's anticipation, as shown in Fig. 8.
4. The accuracy of infant articulation improves along with the separation of his/her prototypes, and the aspect of guidance becomes dominant over the aspect of separation. This trend, where the infant's prototypes are gradually guided toward the anticipated prototypes, can be seen in the last half of the period under conditions (c) and (d) in Fig. 7.

This picture of development includes two new hypotheses to be addressed: 1) It is the inaccuracy of the infant's articulation that separates his/her prototypes, and 2) stretching motherese is a reflection of the caregiver's underestimation of the infant's prototypes expressed through the caregiver's imitation.

We assume that the inaccuracy of infant articulation can be modeled as a Gaussian noise in articulation command space and that its degree is improved along with the separation of the infant's prototypes. However, there seems several possible causes of this inaccuracy, such as the immature levels of three key factors: auditory-articulatory integration, articulatory muscles, and auditory perception. Some studies have addressed issues related to this type of development (Kanda et al., 2008, Guenther and Perkell, 2004, Westermann and Miranda, 2004). Introducing their findings in our model would help us to improve it.

Motherese is one of the well-known characteristics of caregivers addressing infants (Kuhl et al., 1997, Fernald and Simon, 1984), and many researchers have argued for its facilitating role in infant development (Kuhl et al., 2008, Gogate et al., 2006, Liu et al., 2003, Masataka, 1993). However, there have been few explanations proposed for the mechanism behind the rise of motherese. Our results suggest that motherese comes from a caregiver's unconscious anticipation in sharing vowels with her infant in an underestimating manner.

Acknowledgements

This work was supported in part by a Grant-in-Aid for JSPS Fellows.

References

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M., and Yoshida, C. (2009). Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development*, 1(1):12 34.

Bloom, K. and Lo, E. (1990). Adult perceptions of vocalizing infant. *Infant Behavior and Development*, 13:209 219.

de Boer, B. (2000). Self organization in vowel systems. *Journal of Phonetics*, 28(4):441 465.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *J. of the Royal Statistical Society. Series B(Methodological)*, 39:1 38.

Fernald, A. and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20(1):104 113.

Gogate, L. J., Bolzani, L. H., and Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6- to 8-month-old infant and learning of word-object relations. *Infancy*, 9(3):259 288.

Gros-Louis, J., West, M. J., Goldstein, M. H., and King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *Int. J. of Behavioral Development*, 30:509 516.

Guenther, F. H. and Perkell, J. S. (2004). A neural model of speech production and its application to studies of the role of auditory feedback in speech. In *Speech Motor Control in Normal and Disordered Speech*, pages 29 49.

Ishihara, H., Yoshikawa, Y., Miura, K., and Asada, M. (2008). Caregiver's sensorimotor magnets lead infant's vowel acquisition through auto mirroring. In *Proc. of the IEEE 7th Int. Conf. on Development and Learning, CD-ROM.*

Ishizuka, K., Mugitani, R., Kato, H., and Amano, S. (2007). Longitudinal developmental changes in spectral peaks of vowels produced by japanese infant. *Acoustical Society of America*, 121(4):2272 2282.

Kanda, H., Ogata, T., Komatani, K., and Okuno, H. G. (2008). Segmenting acoustic signal with articulatory movement using recurrent neural network for phoneme aquisition. *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 1712 1717.

Kokkinaki, T. and Kugiumutzakis, G. (2000). Basic aspects of vocal imitation in infant-parent interaction during the first 6 monthes. *J. of Reproductive and Infant Psychology*, 18(3):173 187.

Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50:93 107.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhenikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., and Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infant. *Science*, 277:684 686.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). learning as a pathway to language: new data and native language magnet thory expanded. *Philosophical Transactions of the Royal Society B*, 363:979 1000.

Kuhl, P. K. and Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *J. of Acoustic Society of America*, 100:2415 2438.

Liu, H.-M., Kuhl, P. K., and Tsao, F.-M. (2003). An association between mothers' speech clarity and infant' speech discrimination skills. *Developmental Science*, 6:3:F1 F10.

Masataka, N. (1993). Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three- to four-month-old japanese infants. *J. of Child Language*, 20:303 312.

Masur, E. F. and Olson, J. (2008). Mothers' and infant' responses to their partners' spontaneous action and vocal/verbal imitation. *Infant Behavior and Development*, 31:704 715.

McMurray, B., Aslin, R. N., and Tascano, J. C. (2009). Computational principles of language acquishition. *Developmental Science*, 12(3):369 378.

Miura, K., Asada, M., and Yoshikawa, Y. (2007). Unconscious anchoring in maternal imitation that helps finding the correspondence of caregiver's vowel categories. *Advanced Robotics*, 21:1583 1600.

Moody, J. and Darken, C. (1989). Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1(2):281 294.

Oller, D. K. and Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, 59:441 449.

Oudeyer, P.-Y. (2005). The self-organization of speech sounds. *J. of Theoretical Biology*, 233(3):435 449.

Rvachew, S., Mattock, K., and Polka, L. (2006). Developmental and cross-linguistic varidation in the infant vowel space: The case of canadian english and canadian french. *J. of Acoustical Society of America*, 120(4):2250 2259.

Sato, M. and Ishii, S. (2000). On-line EM algorithm for the normalized gaussian network. *Neural Computation*, 12:407 432.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., and Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proc. of National Academy of Sciences USA*, 104(33):13273 13278.

van Beinum, F. J. K., Clement, C. J., and van den Dikkenberg-Pot, I. (2001). Babbling and the lack of auditory speech perception: a matter of coordination? *Developmental Science*, 4:1:61 70.

Vihman, M. M. and Nakai, S. (2003). Experimanhtal evidence for an effect of vocal experience on infant speech perception. *Proc. of the 15 th Int. Cong. of Phonetic Science*, pages 1017 1020.

Vorperian, H. K. and Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *J. of Speech, Language, and Hearing Research*, 50:1510 1545.

Westermann, G. and Miranda, E. R. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language*, 89(2):393 400.