How Caregiver's Anticipation Shapes Infant's Vowel Through Mutual Imitation

Hisashi Ishihara, Yuichiro Yoshikawa, Katsushi Miura, and Minoru Asada, *Fellow, IEEE*

***Abstract*—The mechanism of infant vowel development is a fundamental issue of human cognitive development that includes perceptual and behavioral development. This paper models the mechanism of imitation underlying caregiver–infant interaction by focusing on potential roles of the caregiver's imitation in guiding infant vowel development. Proposed imitation mechanism is constructed with two kinds of the caregiver's possible biases in mind. The first is what we call "sensorimotor magnets," in which a caregiver perceives and imitates infant vocalizations as more prototypical ones as mother-tongue vowels. The second is based on what we call "automirroring bias," by which the heard vowel is much closer to the expected vowel because of the anticipation being imitated. Computer simulation results of caregiver–infant interaction show the sensorimotor magnets help form small clusters and the automirroring bias shapes these clusters to become clearer vowels in association with the sensorimotor magnets.**

***Index Terms*—Caregiver's anticipation , sensorimotor mapping, vowel development.**

I. Introduction

HOW DO infants acquire language? Recognizing and producing voices seem to be a first developmental step for language acquisition. Infants' ability to listen to adult voices appears in a language-independent manner from birth and gradually adapts to their mother tongue [1], [2]. Infant utterances, which are initially quasivocalic sounds that resemble vowels, are gradually adapted to their caregiver ones [3] along with the descent of the epiglottis [4]. Therefore, it seems likely that vocal interaction with their caregivers is needed for infants to adapt their vocal system to their caregivers' language. Some researchers have reported mother characteristic behaviors that seem important for infant vocal development such as infant-directed speech [5]–[9] and imitative interactions [10]–[12]. However, how such caregiver behavior affects infant vocal learning remains unclear due to the difficulties of conducting controlled experiments to understand interaction dynamics.

Manuscript received December 19, 2008; revised November 03, 2009. First published December 18, 2009; current version published February 05, 2010.

H. Ishihara is with the Graduate School of Engineering, Osaka University, Osaka, 565-0871, Japan (e-mail: hisashi.ishihara@ams.eng.osaka-u.ac.jp).

Y. Yoshikawa is with the Asada Synergistic Intelligence Project, ERATO, JST, Osaka, 565-0871, Japan (e-mail: yoshikawa@jeap.org).

K. Miura and M. Asada are with the Graduate School of Engineering, Osaka University, Osaka, 565-0871 Japan, and the Asada Synergistic Intelligence Project, ERATO, JST, Osaka, 565-0871, Japan (e-mail: miura@jeap.org; asada@ams.eng.osaka-u.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAMD.2009.2038988

In recent years, some synthetic studies have been offered as promising approach for modeling such as a dynamic process in the developmental process of infants [13]–[17]. Some of these studies have focused on the perceptual development needed to learn a caregiver's vowel categories from her speech [18]–[20]. However, infants' development of vocal production is also considered in the context of vowel development [21]. The papers by de Boer [22] and Oudeyer [23] focused on imitative interaction and showed that a population of learning agents with a vocal tract and cochlea can self-organize shared vowels among a population through imitative interaction. They assumed that all agents have the same articulation ability, however, an infant and a caregiver cannot perfectly reproduce each other's voices because the infants' articulation abilities have not matured [24], [25]. In other words, they paid less attention to such a specific problem in caregiver–infant interaction, that is to find the phonological correspondence of utterance between caregivers and infants.

On the other hand, Yoshikawa *et al.* [26] addressed this issue in human–robot vocal interaction and demonstrated the importance of imitation by the human caregiver, whose body is different from the robot's. In a similar experimental setting, Miura *et al.* [27] argued that caregiver imitations of infant utterances have two roles: informing the phonological correspondence of utterances and guiding the infant to articulate utterances that more closely resemble the usual ones of the caregiver. The latter is conjectured from the hypothesized characteristics of maternal imitation where a caregiver cannot imitate her infant's behavior due to the embodiment differences; as a result, the presumably imitated behavior is more or less replaced by her accustomed one. Part of this characteristic seems to originate from such caregiver sensorimotor bias as Kuhl's perceptual magnet effect [28], which previous work has already focused on [23]. The perceptual magnet effect indicates a psychological phenomenon where a person recognizes phonemes as more typical ones in the phoneme categories of the person. We call this bias "sensorimotor magnets," conjecturing that our motion also tends to be attracted to the accustomed one. On the other hand, human perception is biased toward anticipation (e.g., [29]), and so is human imitation. It is likely that, in imitative interactions, a caregiver anticipates being imitated by her infant and therefore, perceives the infant's voice as more closely resembling her precedent utterances. We call this the "automirroring bias."

In this paper, we propose a computational model of vowel acquisition through mutual imitation that considers both these biases as causes of the two possible roles of caregiver's imitation: informing of vowel correspondence and guiding infant's vowels to clearer ones. The rest of this paper is structured as

follows. We first propose an imitation mechanism that considers the "automirroring bias" and "sensorimotor magnets," as well as a learning method by mutual imitation with a caregiver. Then a computer simulation of caregiver–infant vowel mutual imitation illustrates how caregivers' sensorimotor magnets help the infant form smaller clusters of vowels and how automirroring bias shapes these clusters to become clear vowels in association with the sensorimotor magnets. Finally, we discuss future issues and conclude the paper.

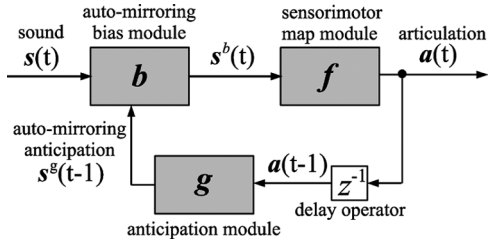


Fig. 1. Proposed imitation mechanism considering biasing elements.

II. Assumption

We assume the following.

1) *Iteration of Multiple Imitative Turn-Takings*
An infant and her caregiver iterate vocal imitative turn-takings. Imitation is considered as one of the mechanism of vocal development [12], [30], and it has been reported that caregiver–infant interaction involves imitative turn-takings [10], [31]. However, prelinguistic infants seldom show vocal imitation in the sense of behavioral matching [32], [33]. Therefore, we assume that infants "try to" reproduce a heard sound while they can not be correctly matched with the target sound.
2) *Initial Categorization of Caregiver Vowels*
Infants have established a rough categorization of caregiver's vowels before simulated caregiver–infant interaction in this study. Observations of early sensitivities for adult vowels [34], [35] support this assumption.
3) *Statistical Learning Capability of Contingent Relation*
An infant statistically learns the correspondence between her articulation and subsequent caregiver's voice. Corresponding to this assumption, three- or four-month-olds can attend to the contingent caregiver's feedback [36], [37], and use it to facilitate development in vocal behavior [36], [38]. Accordingly, infants can rapidly adapt their vocalizations based on the forms of distribution of the voices they have heard [39].
4) *Formant Extractor*
From the heard voices of others, an infant and her caregiver can extract the formant, which is the frequency of peaks that appear in the spectral envelopes of sounds. The lowest two peak frequencies called the first and second formants are used to distinguish vowels by adults [40], [41], and by prelinguistic infants as well [2].
5) *Established Articulation Skills*
Infants have established articulation skills, that is, knowing the connection between an articulatory movement and the sound produced by the movement. However, such a sensorimotor mapping is considered to be developed through vocalizing experiences [42], and previous synthetic studies have addressed this issue [43]–[45]. Although this study ignores it for simplicity, it should be treated as a parallel process of the correspondence of learning vowels through mutual imitation in the future.
6) *Caregiver's Consistent Imitation With Sensorimotor Magnets*
A caregiver can consistently imitate her infant's voice, that is, transforming the perceived voice to its articulation, during the simulated interaction. Furthermore, it is conjectured that the caregiver's imitation is biased to her accustomed vowel utterances due to a sensorimotor magnet extended from Kuhl's perceptual magnet effect [28].
7) *Caregiver's Automirroring Bias*
Human perception is biased toward anticipation (e.g., [29]), and caregivers are likely to anticipate being imitated by infants in imitative interactions. Therefore we assume that the caregivers perceives infant's voices as more closely resembling their precedent utterances (automirroring bias).
8) *Infant's Unexpressed Biases*
It is reported that infants begin to show the perceptual magnet effect from the time they are six months old [7], and develop social expectancy by three months of age [46]. To focus on the role of the caregiver's biases in the current paper, we ignore these biases of the infant as they are considered to be still weak in the early stage of the development.

III. Imitation Mechanism Considering Biasing Elements

Suppose that two people alternately iterate and imitate each other's voices (see assumption 1), and that the sound can be denoted by an N_s dimensional vector and the articulation to produce the imitation sound can be denoted by an N_a dimensional vector.

Fig. 1 illustrates the imitation process by the proposed mechanism: at the tth step of mutual imitation, it listens to the other's voice $\boldsymbol{s}(t) \in \Re^{N_s}$ and imitates $\boldsymbol{s}(t)$ by articulation $\boldsymbol{a}(t) \in \Re^{N_a}$. This imitation process consists of three functions: an automirroring bias module that biases input sounds, a sensorimotor map module that produces an imitation utterance from biased input, and an anticipation module that calculates what we call "automirroring anticipation" from one's last imitation utterance. "Automirroring anticipation" is defined as the perceptual bias by which other's voices are heard as if they resemble the listener's own precedent utterances because of the listener's anticipation being imitated. In the tth step of the imitation trials, the other's heard voice $\boldsymbol{s}(t) \in \Re^{N_s}$ is input to automirroring bias module $\boldsymbol{b} : \Re^{N_s} \rightarrow \Re^{N_s}$, which attracts $\boldsymbol{s}(t)$ to automirroring anticipation $\boldsymbol{s}^g(t-1) \in \Re^{N_s}$. This biased sound $\boldsymbol{s}^b(t)$ is input to the sensorimotor map module and converted to articulation $\boldsymbol{a}(t)$ by function $\boldsymbol{f} : \Re^{N_s} \rightarrow \Re^{N_a}$. $\boldsymbol{a}(t)$ is an imitation utterance of $\boldsymbol{s}(t)$. Moreover, imitation utterance $\boldsymbol{a}(t)$ is input to the anticipation module and converted to automirroring anticipation $\boldsymbol{s}^g(t)$ by function $\boldsymbol{g} : \Re^{N_a} \rightarrow \Re^{N_s}$. Automirroring anticipation $\boldsymbol{s}^g(t)$

is input to the automirroring bias module as an attractor for the other's next voice $\boldsymbol{s}(t+1)$.

A. Automirroring Bias Module

Other's voice $\boldsymbol{s}(t)$ is biased to automirroring anticipation $\boldsymbol{s}^g(t-1)$ and converted to $\boldsymbol{s}^b(t)$ that is given by

$$\begin{aligned}\boldsymbol{s}^b(t) &= \boldsymbol{b}\left(\boldsymbol{s}(t), \boldsymbol{s}^g(t-1); \alpha\right) \\ &= \boldsymbol{s}(t) + \alpha\left(\boldsymbol{s}^g(t-1) - \boldsymbol{s}(t)\right) \quad (0.0 \le \alpha \le 1.0) \end{aligned} \tag{1}$$

where α is a parameter that determines the strength of the automirroring bias (see assumption 7). When α is close to 0, output $\boldsymbol{s}^b(t)$ nearly equals original input $\boldsymbol{s}(t)$ since the automirroring bias is weak. Conversely, when α is close to 1, output $\boldsymbol{s}^b(t)$ is almost attracted to automirroring anticipation $\boldsymbol{s}^g(t-1)$.

B. Sensorimotor Map Module

Since human adults and infants (and robots also) do not have completely identical sensorimotor systems, they cannot perfectly reproduce the other's voices. Therefore, these other voices need to be converted into articulation parameters that generate the listener's own utterable vowels. We use the normalized Gaussian network (NGnet) to map the other's utterable vowel region onto the listener's own generable articulation parameter space (see assumption 6). NGnet is a modular probabilistic regression function that maps N_s-dimensional input space onto N_a-dimensional output space with M units. NGnet $\boldsymbol{f}$ is defined by

$$\boldsymbol{a}(t) = \boldsymbol{f}\left(\boldsymbol{s}^b(t); \theta^f\right) = \sum_{i=1}^{M} \mathcal{N}_i\left(\boldsymbol{s}^b(t)\right) \tilde{\boldsymbol{W}}_i^f \tilde{\boldsymbol{s}}(t) \tag{2}$$

where $\tilde{\boldsymbol{s}}^b$ is the augmented vector of $\boldsymbol{s}^b$ and $\left(\tilde{\boldsymbol{s}}^b\right)^{\mathrm{T}} \equiv \left(\left(\boldsymbol{s}^b\right)^{\mathrm{T}}, 1\right)$. Moreover, $\tilde{\boldsymbol{W}}_i^f \in \Re^{N_a \times (N_s+1)} \equiv (\boldsymbol{W}_i^f, \boldsymbol{r}_i)$ and $\boldsymbol{W}_i^f$ are linear regression matrices. $\mathcal{N}_i(\boldsymbol{s}^b(t))$ is a ith normalized Gaussian function such as

$$\mathcal{N}_i\left(\boldsymbol{s}^b(t)\right) \equiv G_i\left(\boldsymbol{s}^b(t)\right) / \sum_{j=1}^{M} G_j\left(\boldsymbol{s}^b(t)\right) \tag{3}$$

where G_i is a Gaussian function whose center is $\boldsymbol{\mu}_i^f \in \Re^{N_s}$ and whose covariance matrix is $\boldsymbol{\Sigma}_i^f \in \Re^{N_s \times N_s}$, such as

$$\begin{aligned} G_i\left(\boldsymbol{s}^b(t)\right) \equiv{}& (2\pi)^{-\frac{N}{2}} \left|\boldsymbol{\Sigma}_i^f\right|^{-\frac{1}{2}} \\ &\times \exp\left[-\frac{1}{2}\left(\boldsymbol{s}^b(t) - \boldsymbol{\mu}_i^f\right)^{\mathrm{T}} \left(\boldsymbol{\Sigma}_i^f\right)^{-1} \left(\boldsymbol{s}^b(t) - \boldsymbol{\mu}_i^f\right)\right] \end{aligned} \tag{4}$$

where $|\boldsymbol{\Sigma}_i^f|$ is a determinant of matrix $\boldsymbol{\Sigma}_i^f$. Note that we denote a set of parameters of an NGnet $\boldsymbol{f}\{\boldsymbol{\mu}_i^f, \boldsymbol{\Sigma}_i^f, \tilde{\boldsymbol{W}}_i^f | i = 0, \ldots, M\}$ as θ^f.

Normalized Gaussian functions $\mathcal{N}_i(\boldsymbol{s}^b(t))(i = 1, \ldots, M)$ moderately partition the input space into M regions. The ith unit linearly approximates its output by $\tilde{\tilde{\boldsymbol{W}}}_i^f \tilde{\boldsymbol{s}}(t)$ within the corresponding region. NGnet output is given by a summation of these outputs weighted by the normalized Gaussian functions, as in (2).

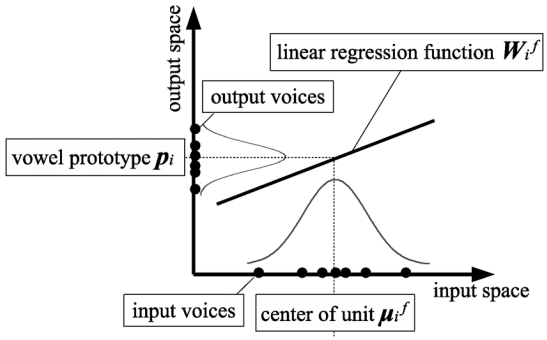


Fig. 2. Illustration of sensorimotor magnets with linear regression function.

Sensorimotor magnets are represented by NGnet $\boldsymbol{f}$ in this module. Fig. 2 shows how sensorimotor magnets are illustrated where we suppose that input data are normally distributed with a central focus on the center of an NGnet unit. The distribution of output data belong to the ith unit is determined by the linear regression matrix $\boldsymbol{W}_i^f$ of the NGnet $\boldsymbol{f}$. Note that $\boldsymbol{W}_i^f$ can control the variance of them: the smaller the eigenvalues of $\boldsymbol{W}_i^f$ are, the more the distribution shrinks to $\tilde{\boldsymbol{W}}_i^f \tilde{\boldsymbol{\mu}}_i^f$, where $\left(\tilde{\boldsymbol{\mu}}_i^f\right)^{\mathrm{T}} \equiv \left(\left(\boldsymbol{\mu}_i^f\right)^{\mathrm{T}}, 1\right)$. Assuming to use such $\boldsymbol{W}_i^f$, we call the center of the transfered voices $\tilde{\boldsymbol{W}}_i^f \tilde{\boldsymbol{\mu}}_i^f$ the ith vowel prototype p_i.

C. Anticipation Module

This module converts articulation $\boldsymbol{a}(t-1)$ to automirroring anticipation $\boldsymbol{s}^g(t-1)$. We use NGnet $\boldsymbol{g}$ to map N_a-dimensional input space onto N_s-dimensional output space contrary to NGnet $\boldsymbol{f}$ in the sensorimotor map module. Automirroring anticipation is calculated by

$$\boldsymbol{s}^g(t-1) = \boldsymbol{g}\left(\boldsymbol{a}(t-1); \theta^g\right) \tag{5}$$

where $\theta^g \equiv \{\boldsymbol{\mu}_i^g, \boldsymbol{\Sigma}_i^g, \tilde{\boldsymbol{W}}_i^g | i = 0, \ldots, M\}$ is a set of parameters of NGnet $\boldsymbol{g}$.

IV. Learning Method for Infants

We assume that a simulated infant (hereinafter infant) initially has an immature imitation mechanism; the parameters of NGnet $\boldsymbol{f}$, i.e., θ^f, in the sensorimotor map module are estimated through mutual imitation. Before the learning, in other words, her vowel prototypes $\boldsymbol{p}_i$ are not clear vowels by which she cannot accurately imitate utterances of a simulated caregiver (hereinafter caregiver). Furthermore, we assume that she does not have automirroring bias, i.e., $\alpha = 0$, for the simplicity of the first simulation trial (see assumption 8). Here the infant's task is tuning parameters θ^f to match vowel prototypes $\boldsymbol{p}_i$ with the clearest vowels for a caregiver by mutual imitation.

In the Tth step of the imitation trials, an infant utters articulation $\boldsymbol{y}(T)$, and a caregiver utters $\boldsymbol{x}(T)$, which imitates $\boldsymbol{y}(T)$.

The infant updates parameters θ^f with the EM algorithm for the NGnet [47], using the caregiver's voice at the last n steps $\boldsymbol{x}(t)(t = T - n + 1, \ldots, T)$ as input data and her own utterances at the last n steps $\boldsymbol{y}(t)(t = T - n + 1, \ldots, T)$ as output data (see assumption 3).

V. Simulation of Vowel Mutual Imitation

We investigate the effects of the biasing elements on vowel learning by simulating the caregiver–infant mutual imitation of vowels with two imitation mechanisms.

A. Procedure

In the simulations, an infant and a caregiver alternately imitate one another with their imitation mechanisms (see assumption 1). The infant has an immature imitation mechanism and updates parameters θ^f of NGnet $\boldsymbol{g}$ with our proposed learning method, and the caregiver has a mature imitation mechanism, so her imitation parameters are fixed during a session of iterating mutual imitations (see assumption 6).

A caregiver imitates her infant's voice every step. Until n steps have passed, the infant selects voices randomly with normal distributions whose centers are her initial vowel prototypes and utters them. After the nth step, the infant basically imitates the caregiver's voice every step, but every fifth step she randomly selects voices with normal distributions whose centers are her current vowel prototypes and utters them. Until n steps have passed, the infant does not update imitation parameters θ^f since she does not have enough learning data. We determined her initial imitation parameters so that her initial vowel prototypes are randomly located in her vowel region while the initial centers of Gaussian function $\boldsymbol{\mu}_i^f$ in her sensorimotor map module are randomly located around a caregiver's vowel prototypes (see assumption 2). In the simulations, $n = 500$ and total learning steps $T_L = 5000$.

B. Settings

We determined each utterable vowel region and the locations of the caregiver's vowel prototypes by imagining a real caregiver and an infant. Fig. 3 shows the vowel region of the real infants and the adults [3], [25]. Vowel prototypes are distinguishable in 2-dimensional vowel space, which is represented by the first (F1) and the second formant frequencies (F2). As shown in Fig. 3, the vowel regions of the real infants and adults are different. For the current simulation, the vowel regions both of the caregiver and the infant are determined in 2-dimensional vowel space, as shown in Fig. 4 (see assumption 4), so the differences between the caregivers and the infants are highlighted. We also regard this vowel space as the articulation parameter space, that is, the vowels and articulation parameters to generate these vowels are the same 2-dimensional vector (see assumption 5).

C. Mature Imitation Mechanism for Caregiver

We determined the locations of caregiver vowel prototypes $\boldsymbol{p}_i^c(i = 1, \ldots, M)$ and their number M, assuming that she uses the five Japanese vowels in the simulation. Therefore, as shown in Fig. 4, the number of vowel prototypes, that is, the number of units M of NGnet f^c in the caregiver's sensorimotor map module, is set to five. Note that super suffix "c" indicates the caregiver's imitation parameters. Furthermore, we assume that a caregiver knows the clearest vowels $\boldsymbol{p}_i^{\text{target}}(i = 1, \ldots, M)$ in an infant vowel region; we determined these clearest vowels as

$$\boldsymbol{p}_i^{\text{target}} = \boldsymbol{p}_i^c + \begin{pmatrix} 400 \\ 600 \end{pmatrix} \tag{6}$$

where $\boldsymbol{p}_i^c = \tilde{\boldsymbol{W}}_i^{fc} \tilde{\boldsymbol{\mu}}_i^{fc}$, which are the vowel prototypes of the caregiver. In the simulations, clearest vowels $\boldsymbol{p}_i^{\text{target}}$ are the target vowels for an infant; in other words, the task is to match her vowel prototypes $\boldsymbol{p}_i \equiv \tilde{\boldsymbol{W}}_i^f \tilde{\boldsymbol{\mu}}_i^f$ with $\boldsymbol{p}_i^{\text{target}}$, that are the clearest vowels in an infant vowel region for her mother.

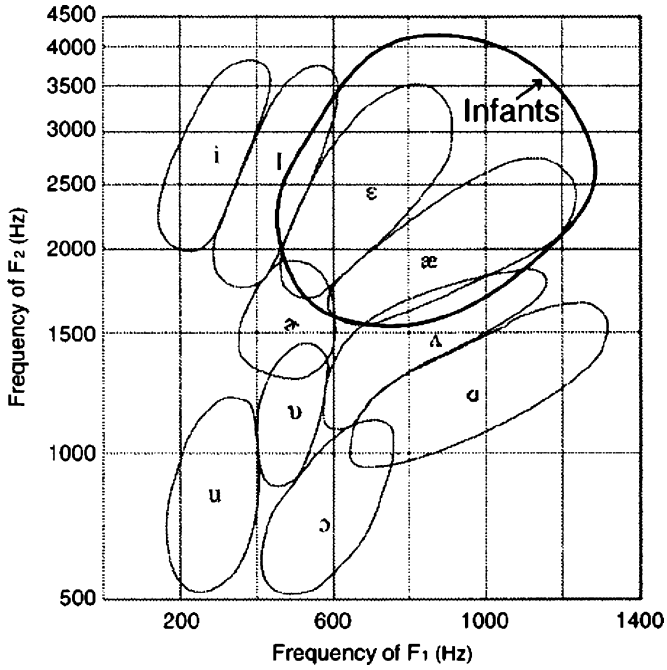


Fig. 3. Vowel regions of real adults and infants in 2-dimensional formant space (Kuhl's plot [3] of infant vowel region in relation to the plot published by Peterson and Barney [25] based on vowel productions of 76 men, women, and children).

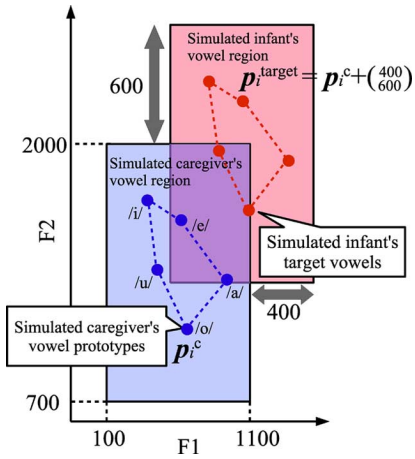


Fig. 4. Settings of two vowel regions of simulated infant and caregiver.

Considering all of the above assumptions, we determined the parameters of NGnet f^c in a caregiver's sensorimotor map module as the following

$$\boldsymbol{\mu}_i^{fc} = \boldsymbol{p}_i^c + \begin{pmatrix} 400 \\ 600 \end{pmatrix} \quad (i = 1, \ldots, M) \tag{7}$$

$$\boldsymbol{\Sigma}_i^{fc} = \begin{pmatrix} 3600 & 0 \\ 0 & 3600 \end{pmatrix} \quad (i = 1, \ldots, M) \tag{8}$$

$$\tilde{\boldsymbol{W}}_i^{fc} = \left((1-\beta^c)\boldsymbol{I}, \boldsymbol{p}_i^c - (1-\beta^c)\boldsymbol{\mu}_i^{fc} \right) \quad (i = 1, \ldots, M, 0.0 \leq \beta^c < 1.0) \tag{9}$$

where β^c is a parameter that determines the strength of the sensorimotor magnets. When β^c is close to 0(1), a caregiver's imitation voice corresponds almost exactly to the infant utterances (either of her vowel prototypes).

In addition, we determined the parameters of NGnet $\boldsymbol{g}^c$ in the caregiver's anticipation module as follows

$$\boldsymbol{\mu}_i^{gc} = \boldsymbol{p}_i^c \quad (i = 1, \ldots, M) \tag{10}$$

$$\boldsymbol{\Sigma}_i^{gc} = \begin{pmatrix} 3600 & 0 \\ 0 & 3600 \end{pmatrix} \quad (i = 1, \ldots, M) \tag{11}$$

$$\tilde{\boldsymbol{W}}_i^{gc} = (\boldsymbol{I}, \boldsymbol{\mu}_i^{gc} - \boldsymbol{p}_i^c) \quad (i = 1, \ldots, M). \tag{12}$$

The caregiver's imitation mechanism has two parameters that determine the strength of the biasing elements: α^c for the automirroring bias, and β^c for the sensorimotor magnets. We investigated the effects of the biasing elements on the learning result of an infant by simulating the interactions and changing these parameters.

D. Immature Imitation Mechanism for Infants

In this study, we assume that an infant initially has rough knowledge about her caregiver's vowel prototypes (see assumption 2), but she cannot know which vowel corresponds to which prototype within her own vowel region. Based on these assumptions, we randomly give initial parameters to the EM algorithm every step as follows

$$\boldsymbol{\mu}_i^f = \mathcal{N}\left(\boldsymbol{p}_i^c, 200^2 I\right) \quad (i = 1, \ldots, M) \tag{13}$$

$$\boldsymbol{\Sigma}_i^f = \begin{pmatrix} \mathcal{N}(3600, 30^2) & 0 \\ 0 & \mathcal{N}(3600, 30^2) \end{pmatrix} \quad (i = 1, \ldots, M) \tag{14}$$

$$\tilde{W}_i^f = \begin{pmatrix} \mathcal{N}(1, 0.5^2) & \mathcal{N}(0, 0.5^2) & \mathcal{N}(500, 500^2) \\ \mathcal{N}(0, 0.5^2) & \mathcal{N}(1, 0.5^2) & \mathcal{N}(500, 500^2) \end{pmatrix} \quad (i = 1, \ldots, M) \tag{15}$$

where $\boldsymbol{p}_i^c$ is the ith vowel prototype of the caregiver and $\mathcal{N}(\boldsymbol{u}, \boldsymbol{v})$ denotes a random value sampled from normal distribution with center $\boldsymbol{u}$ and covariance matrix $\boldsymbol{v}$.

VI. Results

A. Interaction Transitions

Fig. 5 shows the transition of the vowel clarity of the infant voices, the caregiver voices, and the infant vowel prototypes where the caregiver has all the biasing elements ($\alpha^c = 0.5$, $\beta^c = 0.6$). In this graph, the horizontal axis shows the learning steps, and the three curves indicate five times the average of 500 steps' moving average of the following distances: 1) from an infant voice to her nearest target vowel, i.e., clearest vowel in the infant's vowel region; 2) from a caregiver's voice to her nearest vowel prototype, i.e., clearest vowel in the caregiver's vowel region; 3) average distance from each target vowel to her nearest vowel prototype of the infant in each step for evaluating the vowel clarity of each of the above. This graph indicates that although infant voices are not as clear as caregiver voices in the early steps, they became clearer over the time-steps as well as the infant vowel prototypes.

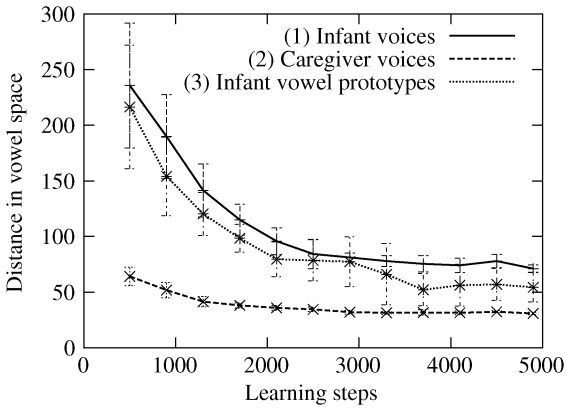


Fig. 5. Transitions of vowel clarity of infant voices, caregiver voices, and infant vowel prototypes where caregiver has all biasing elements ($\alpha^c = 0.5$, $\beta^c = 0.6$).

B. Difference of Learning Results Under Several Conditions

Fig. 6 shows the differences of the learning results under several conditions where the strengths of the caregiver's biasing elements are different and each distribution is an example of the result under each condition. We simulated interaction under the following conditions:

(a) where a caregiver has both automirroring bias and sensorimotor magnets ($\alpha^c = 0.5$, $\beta^c = 0.6$);
(b) where a caregiver only has automirroring bias ($\alpha^c = 0.5$, $\beta^c = 0.0$);
(c) where a caregiver only has sensorimotor magnets ($\alpha^c = 0.0$, $\beta^c = 0.6$);
(d) where a caregiver has no biasing element ($\alpha^c = 0.0, \beta^c = 0.0$).

In these distributions, red (blue) dots represent the infant voices $\boldsymbol{y}(t)$ (the caregiver voices $\boldsymbol{x}(t)$) in the vowel space in the final 1000 steps. The apexes of the red (blue) pentagons represent the target vowels of the infant $\boldsymbol{p}_i^{\text{target}}$ (caregiver vowel prototypes $\boldsymbol{p}_i^c$). Black dots represent the vowel prototypes $\boldsymbol{p}_i$ of the infant after learning. These distributions indicate that the caregiver's biasing elements heavily affected the results of the infant's learning; voice clusters seem smaller under conditions (a) and (c) than under conditions (b) and (d).

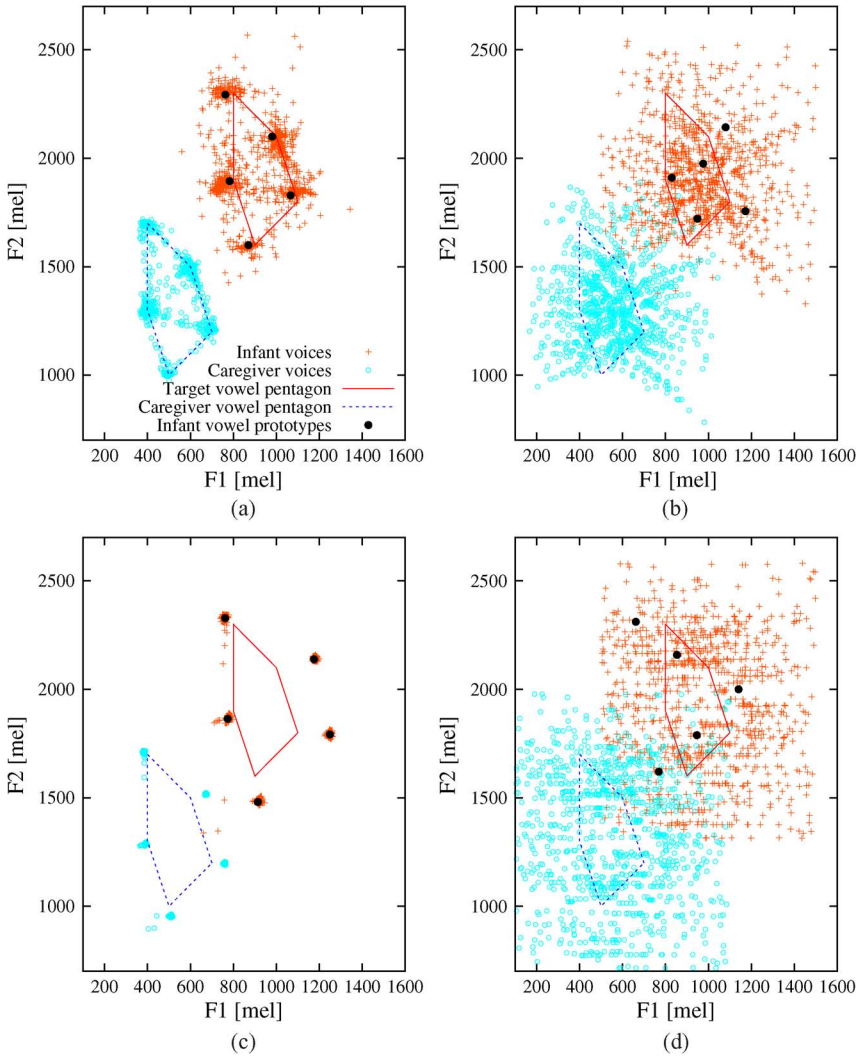


Fig. 6. Difference of learning results under several conditions. Apexes of red pentagons represent target vowels of infant, in other words, clearest vowels in her vowel region, and black dots represent infant vowel prototypes after learning. (a) Both biasing elements; (b) only automirroring bias; (c) only sensorimotor magnets; (d) no biasing elements.

VII. Discussion

A. *Effect of Caregiver's Sensorimotor Magnets*

We can see smaller voice clusters under conditions where a caregiver has sensorimotor magnets in Fig. 6(a) and (c). This suggests that the caregiver's sensorimotor magnets might affect the formation of such voice clusters.

To investigate the relation between the caregiver's sensorimotor magnets and voice cluster formation, we further simulated the interaction under several conditions where the strengths of the caregiver's sensorimotor magnets were different. Fig. 7 shows the relationship between sensorimotor magnet strength β^c (horizontal axis) and the extent of infant voice convergence (vertical axis) during final 1000 steps, which is the averaged value of the five-time simulation with each β^c in three α^c conditions. This figure indicates that the stronger the sensorimotor magnets are, the more tightly the infant voice clusters are gathered and bundled. This would be because the caregiver's imitations form smaller clusters than infant's and accordingly her imitations also form smaller clusters than before.

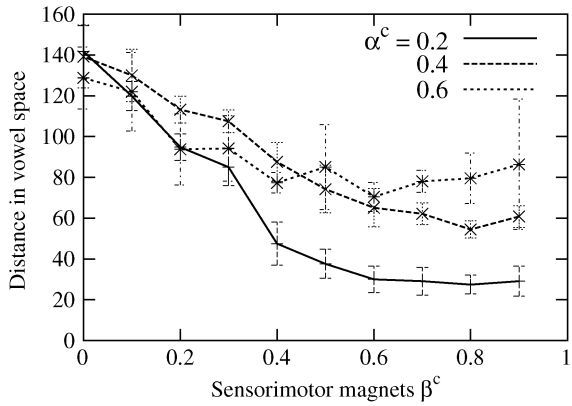


Fig. 7. Different extents of infant voice convergence during final 1000 steps in several conditions.

The centers of these clusters, however, are not always the clearest vowels with this biasing element, as shown in Fig. 6(c).

This can be explained as follows. The caregiver's imitation is shifted to clearer vowels due to her sensorimotor magnets. However, subsequent infant imitation would not be shifted to clearer vowels, that is, would be the almost same as before because she has already tightly associated her caregiver's current voices with her own precedent voice.

B. *Effect of Caregiver's Automirroring Bias*

We refocus on the result shown in Fig. 6. The rating of infant vowel prototypes, expressed by the average distance from each target vowel to the nearest infant vowel prototype, is relatively higher in condition (a) than in condition (c). Although this suggests that the caregiver's automirroring bias helps the infant vowel prototypes approach clearer vowels, this effect probably influenced by the effect of the caregiver's sensorimotor magnets. This is because the rating of the infant's vowel prototypes is lower in condition (b) than in condition (a), although the strengths of the automirroring bias have the same degree.

This can be explained as follows. Although infant's voice is the same as before, her caregiver would perceives it as the clearer vowel than before when she has both biasing elements. This is because her perception is biased to her anticipation and this bias is toward the same direction as her sensorimotor magnets namely clearer vowels. In this way, the caregiver's imitation more or less is considered to be shifted to be clearer vowels than before and accordingly, the infant's imitation would be also shifted to be clearer vowels since the infant's sensorimotor map does not change so drastically bound by previous data. Thus, the infant's voices are considered to be gradually guided to clearer vowels according to her caregiver's anticipation when the caregiver has both biasing elements.

Note that this guidance would emerge only when these biases are not so strong that the infant's sensorimotor map gets heavily inaccurate. To further investigate the effect of caregiver's automirroring bias with the effect of the caregiver's sensorimotor magnets, we simulated interaction in several conditions where the strengths of both automirroring bias and sensorimotor magnets were different. Fig. 8 shows the relationship between the strengths of automirroring bias α^c (vertical axis) and sensorimotor magnets β^c (horizontal axis), and the average distance from each target vowel to the nearest vowel prototypes of the infant after learning in each condition (color map from red to yellow), which is the averaged value of the five-time simulation with each set of α^c and β^c. This figure indicates that the emergence of the guidance requires a balanced association between automirroring bias and sensorimotor magnets.

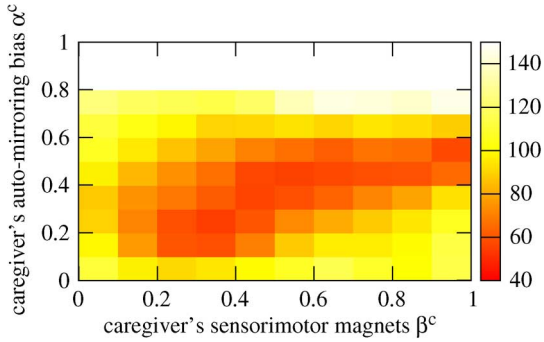


Fig. 8. Difference of ratings of infant vowel prototypes after learning in several conditions.

VIII. Conclusion

We simulated caregiver–infant imitative interaction of vowels, considering the caregiver's biasing elements: automirroring bias and sensorimotor magnets. Simulation results indicate that these biasing elements of the caregiver guide the infant vowel prototypes to become clear vowels; the sensorimotor magnets help form small vowel clusters and the automirroring bias shapes these clusters to become clearer vowels in association with the sensorimotor magnets. The results might imply general importance of the caregiver's anticipation of the infant's ability on guiding various social developments of infants. The current study is limited in the context of imitative interaction of vowels and therefore, applying this developmental mechanism, i.e., guidance by social feedback with the caregiver's anticipation, to other social development is one of our future issues.

In the simulations, we assumed that a caregiver always imitates an infant. However, there exist discussions concerning whether imitation is the responsible mechanism for vocal development [48], and some observational data suggest infant vocal development without a caregiver's imitation [49]. In our model, infants need to be imitated for learning correspondence of utterances between themselves and caregivers. However, real caregivers do not always imitate their infant. We have to extend our model so that we can explain also how caregiver's nonimitative feedback affects to the infant vocal development.

Furthermore, we fixed the strength of the automirroring bias of the caregiver during interactions and assumed that an infant does not have automirroring bias. We will investigate the development of such caregiver and infant parameters. The automirroring bias of a caregiver might become stronger as the infant imitation becomes more accurate since it seems to depend on the extent to which the caregiver anticipates her infant imitation. On the other hand, infants are considered to develop social expectancy of their caregiver's feedback through social interaction [50]. Infant's biased interpretation by their social anticipation, in other words, their automirroring bias, should be considered in the developmental model. In parallel with such synthetic simulations, we have to conduct some psychological experiments to examine characteristics and factors of the caregiver's automirroring bias, for example, what kinds of infants' behavior induce this bias in the caregiver's imitation.

We consider that automirroring bias plays important roles not only in guiding infant vowel prototypes to become clear vowels, but also in maintaining caregiver–infant interaction. We expect that automirroring bias forms an intrapersonal positive feedback loop between the observation to be imitated and the feeling that the opponent is imitating. In their spiral response-cascade hypothesis [51], Yoshikawa *et al.* suggest the existence of inter- and intrapersonal positive feedback loops, not only between observation and feeling, but also between feeling and action. They explain the mechanism responsible for the emergence and maintenance of communication between agents, not just between a caregiver and an infant. Additionally, Stern [52] explains social

interaction and development are made up by caregivers' affect attunement, in which caregivers speculate their infants' mental state, such as intention, and express the mental state they speculate from the infants' behaviors. We believe that automirroring bias is an instance of intrapersonal facilitation on interaction, and imitative interaction continues with the support of the bias. Investigating the maintaining function of automirroring bias is one of important future issues.

REFERENCES

[1] J. F. Werker and R. C. Tees, "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," *Inf. Behav. Develop.*, vol. 25, pp. 121–133, 2002.

[2] B. Cheour, R. Ceponiene, A. Lehtokoski, A. Luuk, J. Allik, K. Alho, and R. Näätänen, "Development of language-specific phoneme representation in the infant brain," *Nature Neurosci.*, vol. 1, pp. 351–353, 1998.

[3] P. K. Kuhl and A. N. Meltzoff, "Infant vocalizations in response to speech: Vocal imitation and developmental change," *J. Acoust. Soc. Amer.*, vol. 100, pp. 2415–2438, 1996.

[4] C. T. Sasaki, P. A. Levine, J. T. Laitman, and E. S. Crelin, "Postnatal developmental descent of the epiglottis in man," *Arch. Otolaryngol.*, vol. 103, pp. 169–171, 1977.

[5] J. F. Werker, F. Pons, C. Dietrich, S. Kajikawa, L. Fais, and S. Amano, "Infant-directed speech supports phonetic category learning in english and japanese," *Cognition*, vol. 103, pp. 147–162, 2007.

[6] P. K. Kuhl, "A new view of language acquisition," in *Proc. Nat. Acad. Sci.*, 2000, vol. 97, no. 22, pp. 11 850–11 857.

[7] P. K. Kuhl, K. A. Williams, F. Lacerda, K. N. Stevens, and B. Lindblom, "Linguistic experience alters phonetic perception in infants by 6 months of age," *Science*, vol. 255, pp. 606–608, 1992.

[8] K. Bloom, A. Russell, and K. Wassenberg, "Turn taking affects the quality of infant vocalizations," *J. Child Lang.*, vol. 14, pp. 211–227, 1987.

[9] K. Bloom, "Quality of adult vocalizations affects the quality of infant vocalizations," *J. Child Lang.*, vol. 15, pp. 469–480, 1987.

[10] T. Kokkinaki and G. Kugiumutzakis, "Basic aspects of vocal imitation in infant-parent interaction during the first 6 months," *J. Reproduct. Inf. Psychol.*, vol. 18, no. 3, pp. 173–187, 2000.

[11] E. F. Masur and J. E. Rodemaker, "Mothers' and infants' spontaneous vocal, verbal, and action imitation during the second year," *Merrill-Palmer Quart.*, vol. 45, no. 3, pp. 392–412, 1999.

[12] M. Papousek and H. Papousek, "Forms of functions of vocal matching in interactions between mothers and their precanonical infant," *First Lang.*, vol. 9, no. 6, pp. 137–157, 1989.

[13] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida, "Cognitive developmental robotics: A survey," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 1, pp. 12–34, May 2009.

[14] A. Stoytchev, "Some basic principles of developmental robotics," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 2, pp. 122–130, Aug. 2009.

[15] J. Weng, "Developmental robotics: Theory and experiments," *Int. J. Human. Robot.*, vol. 1, no. 2, pp. 199–236, 2004.

[16] M. Asada, K. F. MacDorman, H. Ishiguro, and Y. Kuniyoshi, "Cognitive developmental robotics as a new paradigm for the design of humanoid robots," *Robot. Autonom. Syst.*, vol. 37, pp. 185–193, 2001.

[17] L. Steels, "The methodology of the artificial," *Behav. Brain Sci.*, vol. 24, no. 6, pp. 1077–1078, 2001.

[18] B. McMurray, R. N. Aslin, and J. C. Tascano, "Computational principles of language acquishition," *Develop. Sci.*, vol. 12, no. 3, pp. 369–378, 2009.

[19] B. M. Lake, G. K. Vallabha, and J. L. McClelland, "Modeling unsupervised peceptual category learning," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 1, pp. 35–43, May 2009.

[20] G. K. Vallabha, J. L. McClelland, F. Pons, J. F. Werker, and S. Amano, "Unsupervised learning of vowel categories from infant-directed speech," in *Proc. Nat. Acad. Sci. USA*, 2007, vol. 104, no. 33, pp. 13 273–13 278.

[21] P. K. Kuhl, B. T. Conboy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola, and T. Nelson, "Learning as a pathway to language: New data and native language magnet thory expanded," *Philos. Trans. Roy. Soc. B*, vol. 363, pp. 979–1000, 2008.

[22] B. de Boer, "Self organization in vowel systems," *J. Phonet.*, vol. 28, no. 4, pp. 441–465, 2000.

[23] P.-Y. Oudeyer, "The self-organization of speech sounds," *J. Theoret. Biol.*, vol. 233, no. 3, pp. 435–449, 2005.

[24] H. K. Vorperian and R. D. Kent, "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," *J. Speech Hear. Res.*, vol. 50, pp. 1510–1545, 2007.

[25] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *J. Acoust. Soc. Amer.*, vol. 24, pp. 175–184, 1952.

[26] Y. Yoshikawa, J. Koga, M. Asada, and K. Hosoda, "A constructivist approach to infants' vowel acquisition through mother-infant interaction," *Connect. Sci.*, vol. 15, no. 4, pp. 245–258, 2003.

[27] K. Miura, M. Asada, and Y. Yoshikawa, "Unconscious anchoring in maternal imitation that helps finding the correspondence of caregiver's vowel categories," *Adv. Robot.*, vol. 21, pp. 1583–1600, 2007.

[28] P. K. Kuhl, "Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not," *Percept. Psychophys.*, vol. 50, pp. 93–107, 1991.

[29] H. Merckelbach and V. van de Ven, "Another white christmas: Fancy proneness and reports of "hallucinatory experiences" in undergraduate students," *J. Behav. Therapy Exper. Psycol.*, vol. 32, no. 3, pp. 137–144, 2001.

[30] T. Kokkinaki, "A longitudinal, naturalistic and cross-cultural study on emotions in early infant-parent imitative interactions," *Brit. J. Develop. Psychol.*, vol. 21, pp. 243–258, 2003.

[31] J. Gros-Louis, M. J. West, M. H. Goldstein, and A. P. King, "Mothers provide differential feedback to infants' prelinguistic sounds," *Int. J. Behav. Develop.*, vol. 30, pp. 509–516, 2006.

[32] S. S. Jones, "Imitation in infancy," *Psychol. Sci.*, vol. 18, no. 7, pp. 593–599, 2007.

[33] G. Kugiumutzakis, "Intersubjective vocal imitation in early mother-infant interaction," in *New Perspective in Communicative Development*. Evanston, IL: Routleedge, 1993, pp. 23–47.

[34] M. A. Aldridge, R. D. Stillman, and T. G. Bower, "Newborn categorization of vowel-like sounds," *Develop. Sci.*, vol. 4, no. 2, pp. 220–232, 2001.

[35] P. D. Eimas, E. R. Siqueland, P. Jusczyk, and J. Vigorito, "Speech perception in infants," *Science*, vol. 171, pp. 971–974, 1971.

[36] N. Masataka, "Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three- to four-month-old japanese infants," *J. Child Lang.*, vol. 20, pp. 303–312, 1993.

[37] K. Bloom, "Patterning of infant vocal behavior," *J. Exper. Child Psychol.*, vol. 23, pp. 367–377, 1977.

[38] M. H. Goldstein, A. P. King, and M. J. West, "Social interaction shapes babbling: Testing parallels between birdsong and speech," in *Proc. Nat. Acad. Sci. U.S.A*, 2003, vol. 100, no. 13, pp. 8030–8035.

[39] J. Maye, J. F. Werker, and L. Gerken, "Infant sensitivity to distributional information can affect phonetic discrimination," *Cognition*, vol. 82, pp. B101–B111, 2002.

[40] F. W. Ohl and H. Scheich, "Orderly cortical representation of vowels based on formant interaction," in *Proc. Nat. Acad. Sci.*, 1997, vol. 94, pp. 9440–9444.

[41] J. M. Pickett, "Perception of vowels heard in noises spectra," *J. Acoust. Soc. Amer.*, vol. 29, no. 5, pp. 613–620, 1957.

[42] D. K. Oller and R. E. Eilers, "The role of audition in infant babbling," *Child Develop.*, vol. 59, pp. 441–449, 1988.

[43] H. Kanda, T. Ogata, K. Komatani, and H. G. Okuno, "Segmenting acoustic signal with articulatory movement using recurrent neural network for phoneme aquisition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Nice, France, 2008, pp. 1712–1717.

[44] F. H. Guenther and J. S. Perkell, "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*. Oxford, U.K.: Oxford Univ. Press, 2004, pp. 29–49.

[45] G. Westermann and E. R. Miranda, "A new model of sensorimotor coupling in the development of speech," *Brain Lang.*, vol. 89, no. 2, pp. 393–400, 2004.

[46] E. Bertin and T. Striano, "The still-face responce in newborn, 1.5-, and 3-month-old infants," *Inf. Behav. Develop.*, vol. 29, pp. 294–297, 2006.

[47] M. Sato and S. Ishii, "On-line EM algorithm for the normalized gaussian network," *Neural Comput.*, vol. 12, pp. 407–432, 2000.

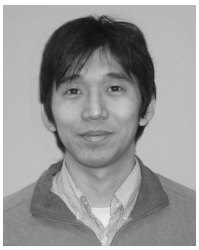
[48] G. Markova and M. Legerstee, "Contingency, imitation, ans affect sharing: Foundations of infants' social awareness," *Develop. Psychol.*, vol. 42, no. 1, pp. 132–141, 2006.

[49] M. H. Goldstein and J. A. Schwade, "Social feedback to infant' babbling facilitates rapid phonological learning," *Psychol. Sci.*, vol. 19, no. 5, pp. 515–523, 2008.

[50] M. Legerstee and J. Varghese, "The role of maternal affect mirroring on social expectancies in three-month-old infant," *Child Develop.*, vol. 72, no. 5, pp. 1301–1313, 2001.
[51] Y. Yoshikawa, S. Yamamoto, H. Sumioka, H. Ishiguro, and M. Asada, "Spiral response-cascade hypothesis -intrapersonal responding-cascade in gaze interaction-," in *Proc. 3rd ACM/IEEE Int. Conf. Human-Robot Interact.*, Amsterdam, The Netherlands, 2008, pp. 319–326.
[52] D. N. Stern, *The Interpersonal World of the Infant*. New York: Basic, 1998.

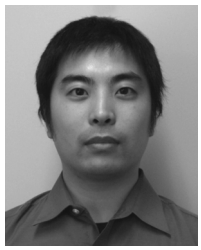
Hisashi Ishihara received the B.Eng. and M.S. degrees in engineering from Osaka University, Osaka, Japan, in 2007 and 2009, respectively. He is currently working towards the Ph.D. degree in the Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University.

From 2009, he has been a Research Fellow of the Japan Society for the Promotion of Science, Tokyo, Japan. His research interests focus on the dynamic structure of social interaction and development in infants.



Yuichiro Yoshikawa received the Ph.D. degree in engineering from Osaka University, Osaka, Japan, in 2005.

From 2003 to 2005, he was a Research Fellow of the Japan Society for the Promotion of Science. From 2005 to 2006, he was a Researcher at Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International. Since April 2006, he has been a Researcher at Asada Synergistic Intelligence Project, ERATO, Japan Science and Technology Agency, Saitama, Japan. He has been engaged in the issues of human–robot interaction and cognitive developmental robotics.



Katsushi Miura received the B.Eng. and M.Eng. degrees in engineering from Osaka University, Osaka, Japan, in 2004, and 2006, respectively.

Since 2006, he has been a Ph.D. candidate of EmergentRobotics Area, Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University and belongs to Asada Synergistic Intelligence Project, ERATO, Japan Science and Technology Agency, Saitama, Japan.

Minoru Asada (F'05) received the B.E., M.E., and Ph.D., degrees in control engineering from Osaka University, Osaka, Japan, in 1977, 1979, and 1982, respectively.

In April 1995, he became a Professor of the Osaka University. Since April 1997, he has been a Professor of the department of Adaptive Machine Systems at the Graduate School of Engineering, Osaka University. From August 1986 to October 1987, he was a Visiting Researcher of Center for Automation Research, University of Maryland, College Park.

Dr. Asada received many awards such as the best paper award of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS92) and the Commendation by the Minister of Education, Culture, Sports, Science and Technology, Japanese Government as Persons of distinguished services to enlightening people on science and technology. He was the president of the International RoboCup Federation (2002–2008). Since 2005, he has been the Research Director of "ASADA Synergistic Intelligence Project" of ERATO (Exploratory Research for Advanced Technology by Japan Science and Technology Agency).