Caregiver's Sensorimotor Magnets Lead Infant's Vowel Acquisition through Auto Mirroring

Hisashi Ishihara[*], Yuichiro Yoshikawa[†], Katsushi Miura[*†], and Minoru Asada[*†]

[*]Graduate School of Eng., Osaka University
2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan
hisashi.ishihara@ams.eng.osaka-u.ac.jp
[†]Asada Synergistic Intelligence Project, ERATO, JST
2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan
{yoshikawa,miura,asada}@jeap.org

***Abstract*—Mother-infant vocal communication is a sort of mystery of human cognitive development since they can communicate although their body structures and therefore their utterable areas are different. This paper proposes a method that aids unconscious guidance in mutual imitation for infant development based on a biasing element with two different kinds of modules. The first is based on the normal magnet effect in perceiving heard vocal sounds as the listener's own vowels (perceptual magnet) and also includes another magnet effect for imitating vocal sounds that resemble the imitator's vowels (articulatory magnet). The second is based on what we call "auto mirroring bias," by which the heard vowel is much closer to the expected vowel because the other's utterance is an imitation of the listener's own utterance. Computer simulation results of mother-infant interaction show the validity of the proposed bias. Finally future issues are discussed.**

I. INTRODUCTION

How can we build robots that can communicate naturally with humans? A bottleneck in such development exists because their body structures differ from humans. On the other hand, the body structure of human infants is immature and not the same as their caregivers, and yet they can acquire similar patterns of communication behavior as their caregivers through interaction with them, e.g., mutual imitation. Modeling the developmental process of infants under such interaction with their caregivers is a formidable challenge. We expect this approach would contribute to building a robot that can acquire human-like communication behaviors through interaction with humans as well as understanding of the cognitive development process of human infants.

Infant studies report that maternal imitation effectively reinforces infant vocalization [1], and infant speech-like cooing tends to elicit utterances from caregivers [2]. Thus mothers/primary caregivers and infants tend to imitate one another's utterances. From the perspective that infants may acquire clear vowels through such mutual imitation, Miura *et al.* experimented on human-robot mutual imitation of vowels and showed that robots which imitated humans voices could enunciate clearer vowels by continuing to adapt their voices to match the caregiver [3]. This implies that human imitation voices include reproduction errors that consequently guide the direction of voices toward clearer ones. Miura *et al.* argued that such maternal guidance resulted from the perceptual magnet effect [4] in which a person mistakenly perceives a sound as being closer to either of her/his own phoneme prototypes than it actually is. The previous work, however, have not considered other causes from the constraints that exist in the human imitation system. Moreover, they gave the robot approximate knowledge about the correspondence relations between human and robot utterable vowels. However, the robot should acquire such knowledge about the correspondence relation through interaction with humans because in most cases providing robots with accurate knowledge about the correspondence relation between humans and robot behavior is difficult.

Keeping these points in mind, we propose a computational model of unconscious guidance in mutual imitation that considers three biasing elements of human imitation. The first is the perceptual magnet effect, as argued in previous work. We call the second an articulatory magnet effect: person's articulation tends to be attracted so that the articulated voice becomes colser to the ones they usualy generate than the one they intended to generate. We believe that both these magnet effects bias the imitation process in the same direction to the person's own primitive voice. Therefore, we model these magnets as a biasing mechanism that we call "sensorimotor magnets" by introducing shrinking properties to a transfer function between the vowel regions of both agents in mutual imitation. Furthermore, we consider a biasing element by analogy with a priming effect, which is a phenomenon that human perception is affected by a prime stimulus. We conjecture that a listener's *prior* utterance works as a prime stimulus in a mutual imitation situation, suggesting that a person tends to perceive another's utterance as being closer to an expected imitation sound of its own *prior* utterance than actual. This is the third biasing element that we call "auto mirroring bias."

The rest of this paper is structured as follows: we first propose an imitation mechanism that considers the auto mirroring bias and sensorimotor magnets as well as a learning method by mutual imitation with a caregiver. Then simulation results about mother-infant vowel mutual imitation illustrate how sensorimotor magnets help learners form smaller clusters of vowels and how the auto mirroring bias guides these clusters to become clear vowels. Finally, we discuss future issues and conclude the paper.

II. Imitation mechanism considering biasing elements

Suppose that two persons alternately iterate and imitate each other's voices and that the sound can be denoted by an N_s dimensional vector while the articulation to produce the imitation sound can be denoted by an N_a dimensional vector.

Figure 1 illustrates the imitation process by the proposed mechanism: at the t-th step of mutual imitation, it listens to the other's voice $\boldsymbol{s}(t) \in \Re^{N_s}$ and imitates $\boldsymbol{s}(t)$ by an articulation $\boldsymbol{a}(t) \in \Re^{N_a}$. It consists of three functions: an auto mirroring biasing module that biases input sounds, a sensorimotor magnet module that produces imitated sound from biased input, and a vowel region mapping module that calculates what we call auto mirroring anticipation from one's last imitation utterance. "Auto mirroring anticipation" is defined as the other person's expected imitation voice of her own present utterance. In the t-th step of the imitation trials, the other's heard voice $\boldsymbol{s}(t) \in \Re^{N_s}$ is input to auto mirroring bias module $\boldsymbol{b} : \Re^{N_s} \to \Re^{N_s}$, which attracts $\boldsymbol{s}(t)$ to auto mirroring anticipation $\boldsymbol{s}^g(t-1) \in \Re^{N_s}$. This biased sound $\boldsymbol{s}^b(t)$ is input to the sensorimotor magnet module and converted to articulation $\boldsymbol{a}(t)$ by function $\boldsymbol{f} : \Re^{N_s} \to \Re^{N_a}$. $\boldsymbol{a}(t)$ is an imitation utterance of $\boldsymbol{s}(t)$. Moreover, imitation utterance $\boldsymbol{a}(t)$ is input to the vowel region mapping module and converted to auto mirroring anticipation $\boldsymbol{s}^g(t)$ by function $\boldsymbol{g} : \Re^{N_a} \to \Re^{N_s}$. Auto mirroring anticipation $\boldsymbol{s}^g(t)$ is input to the auto mirroring bias module as a biasing element for the other's next voice $\boldsymbol{s}(t+1)$.

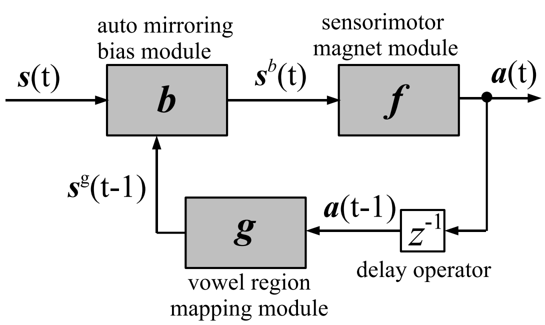


Fig. 1. Proposed imitation mechanism considering biasing elements

A. Auto mirroring bias module

The other's heard voice $\boldsymbol{s}(t)$ is biased to auto mirroring anticipation $\boldsymbol{s}^g(t-1)$ and converted to $\boldsymbol{s}^b(t)$ that is given by

$$\boldsymbol{s}^b(t) = \boldsymbol{b}(\boldsymbol{s}(t), \boldsymbol{s}^g(t-1); \alpha) = \boldsymbol{s}(t) + \alpha(\boldsymbol{s}^g(t-1) - \boldsymbol{s}(t)) \quad (0.0 \leq \alpha \leq 1.0), \quad (1)$$

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

B. Sensorimotor magnet module

Since human adults, infants, and robots do not have completely identical sensorimotor systems, they cannot perfectly reproduce the other's utterances. Therefore the other's voice needs to be converted into the listener's own utterable vowel region. We use the Normalized Gaussian network (NGnet) to map the other's utterable vowel region onto the listener's own utterable vowel region. 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

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

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

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

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

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

where $\left|\boldsymbol{\Sigma}_i^f\right|$ is a determinant of matrix $\boldsymbol{\Sigma}_i^f$. Note that we denote a set of parameters of an NGnet $\boldsymbol{f}$ $\left\{\boldsymbol{\mu}_i^f, \boldsymbol{\Sigma}_i^f, \tilde{\boldsymbol{W}}_i^f | i = 0, \ldots, M\right\}$ 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 i-th unit linearly approximates its output by $\tilde{\boldsymbol{W}}_i^f \tilde{\boldsymbol{s}}(t)$ within the corresponding region. NGnet output is given by a summation of these outputs weighted by the normalized Gaussian functions, as in Eq. (2).

Sensorimotor magnets are represented by NGnet $\boldsymbol{f}$ in this module. Figure 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 is determined by the linear regression matrix of the NGnet. If we regard transferred center $\tilde{\boldsymbol{W}}_i^f \tilde{\boldsymbol{\mu}}_i^f$ as vowel prototypes, where $(\tilde{\boldsymbol{\mu}}_i^f)^{\mathrm{T}} \equiv ((\boldsymbol{\mu}_i^f)^{\mathrm{T}}, 1)$, the linear regression matrix can control the bias strength of the vowel prototype. In other words, the smaller the eigenvalue of $\boldsymbol{W}_i^f$, the more the distribution shrinks.

C. Vowel region mapping module

This module converts the other's utterance $\boldsymbol{a}(t-1)$ to auto mirroring anticipation $\boldsymbol{s}^g(t-1)$. We use NGnet $\boldsymbol{g}$ to map N_a-dimensional input space onto N_s-dimensional output space

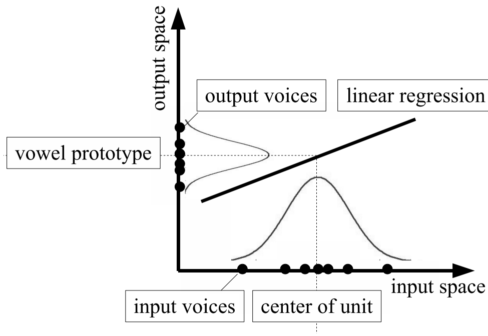


Fig. 2. Illustration of sensorimotor magnets with linear regression function

contrary to NGnet $\boldsymbol{f}$ in the sensorimotor magnet module. Auto mirroring anticipation is calculated by

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

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

III. Learning method for a robot

We assume that the robot initially has an immature imitation mechanism: the parameters of NGnet $\boldsymbol{f}$, i.e., θ^f, in the sensorimotor mapping module are randomly initialized. Furthermore, we assume that it does not have auto mirroring bias, i.e., $\alpha = 0$, for the simplicity of the first simulation trial. As a result, its vowel prototypes are not clear vowels for humans before the learning. Here the robot task is tuning parameters θ^f to match vowel prototype $\tilde{\boldsymbol{W}}_i^f \tilde{\boldsymbol{\mu}}_i^f$ with clear vowels for humans by mutual imitation.

In the T-th step of the imitation trials, a robot utters $\boldsymbol{y}(T)$, a sound in its utterable vowel region, and a caregiver utters $\boldsymbol{x}(T)$, which imitates $\boldsymbol{y}(T)$. The robot updates parameters θ^f with the EM algorithm [5][6][7] using the caregiver's voice at the last n steps $\boldsymbol{x}(t)(t = T-n+1, \ldots, T)$ as input data and its own utterances at the last n steps $\boldsymbol{y}(t)(t = T-n+1, \ldots, T)$ as output data.

IV. Simulation of vowel mutual imitation

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

A. Procedure

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

A mother imitates her infant's voice every step. Until n steps have passed, the infant selects voices randomly with normal distributions whose centers are its initial vowel prototypes and utters them. After the n-th step, the infant basically imitates the mother's voice every step, but every fifth step it randomly selects one of its own vowel prototypes and utters it. Until n steps have passed, the infant does not update imitation parameters θ^f since it does not have enough learning data. We determined its initial imitation parameters so that its initial vowel prototypes are randomly located in its vowel region. In simulations, $n = 500$ and total learning steps $T_L = 5000$.

B. Settings

We determined each utterable vowel region and the locations of the mother's vowel prototypes by imagining a real mother and infant. Figure 3 shows the vowel region of real infants and adults [8][9]. Vowel prototypes are distinguishable in 2-dimensional vowel space, which is represented with the first formant frequency (F1) and the second (F2). As shown in Fig. 3, the vowel regions of real infants and adults are different from each other. For the current simulataion, vowel regions both of the caregiver and the infant are determined in 2-dimensional vowel space as shown in Fig. 4 so that the difference between mother's and infant's one are highlighted.

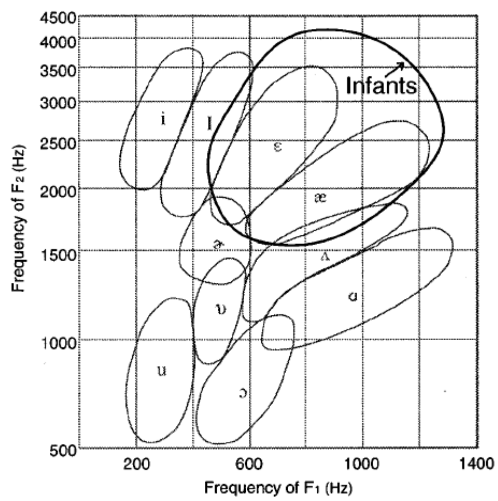


Fig. 3. Vowel regions of real adults and infants in 2-dimensional formant space [8][9]

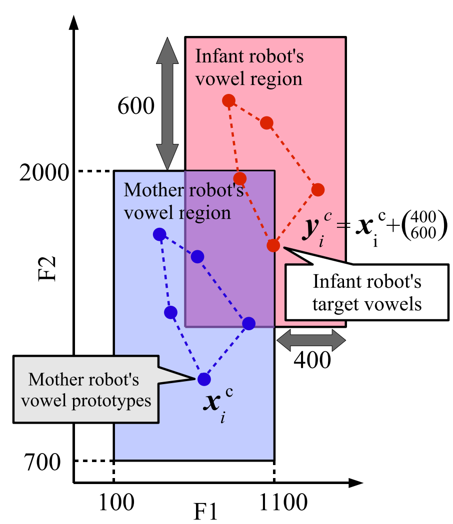


Fig. 4. Settings of two vowel regions of infant and mother robots

C. Mature imitation mechanism for mother

We determined the locations of the mother's vowel prototypes $\boldsymbol{x}_i^c$ and their number, 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 mother's sensorimotor mapping module, is set to five. Note that super suffix 'c' indicates the mother's (**c**aregiver's) imitation parameters. Furthermore, we assume that a mother knows the clearest vowels $\boldsymbol{y}_i^c$ in an infant vowel region; we determined these clearest vowels as

$$\boldsymbol{y}_i^c = \boldsymbol{x}_i^c + \begin{pmatrix} 400 \\ 600 \end{pmatrix}, \tag{6}$$

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

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

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

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

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

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

In addition, we determined the parameters of NGnet $\boldsymbol{g}^c$ in the mother's vowel region mapping module as follows:

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

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

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

The mother's imitation mechanism has two parameters that determine the strength of the biasing elements: one is α^c for auto mirroring bias, and the other is β^c for 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 infant

In this study, we assume that an infant initially does not have accurate knowledge about its mother's vowel prototypes, so it cannot know which vowel corresponds to each of them within its own vowel region. Based on these assumptions, we randomly give initial parameters to the EM algorithm every step as follows:

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

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

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

where $\boldsymbol{x}_i^c$ is the i-th vowel prototype of the mother and $\mathcal{N}(\boldsymbol{p}, \boldsymbol{q})$ denotes a random value sampled from normal distribution with center $\boldsymbol{p}$ and covariance matrix $\boldsymbol{q}$.

V. Results

A. Interaction Transitions

Figure 5 shows the transition of the vowel clarity of the infant utterances, the mother utterances, and the infant vowel prototypes where the mother 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 utterance to its nearest target vowel, i.e., clearest vowel in the infant's vowel region; (2) from a mother utterance to its nearest vowel prototype, i.e., clearest vowel in the mother's vowel region; (3) average distance from each target vowel to its 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 utterances are not as clear as mother utterances in the early steps, they became clearer over the time-steps as well as the infant vowel prototypes.

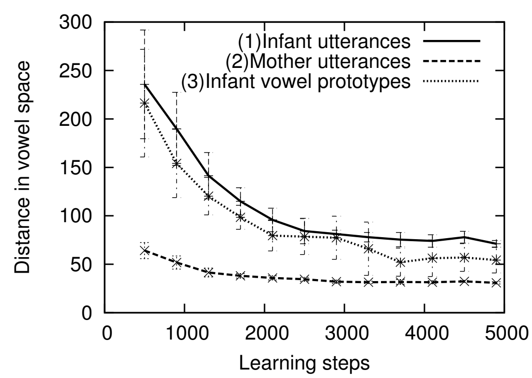


Fig. 5. Transitions of vowel clarity of infant utterances, mother utterances, and infant vowel prototypes where the mother has all biasing elements ($\alpha^c = 0.5, \beta^c = 0.6$).

B. Difference of learning results under several conditions

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

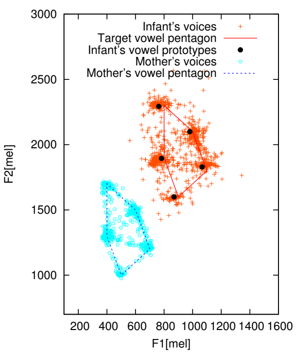
(a) where a mother has both auto mirroring bias and sensorimotor magnets ($\alpha^c = 0.5, \beta^c = 0.6$),

(b) where a mother only has auto mirroring bias ($\alpha^c = 0.5, \beta^c = 0.0$),

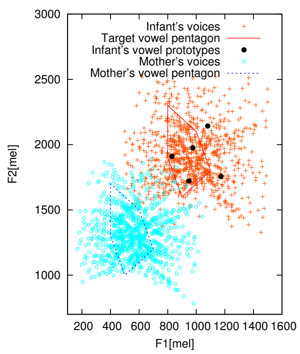
(c) where a mother only has sensorimotor magnets ($\alpha^c = 0.0, \beta^c = 0.6$),

(d) where a mother has no biasing element ($\alpha^c = 0.0, \beta^c = 0.0$).

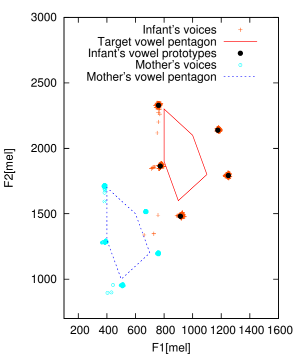
In these distributions, red (blue) dots represent the infant (mother) utterances in the vowel space in the final 1000 steps. The apexes of the red (blue) pentagons represent the target vowels of the infant (mother vowel prototypes). Black dots represent the vowel prototypes of the infant after learning. These distributions indicate that the mother'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).



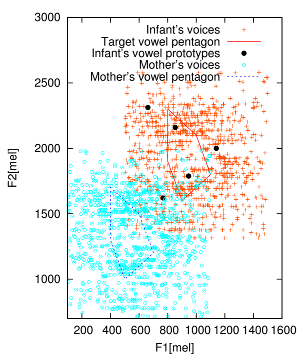
(a) Both biasing elements



(b) Only auto mirroring bias



(c) Only sensorimotor magnets



(d) No biasing element

Fig. 6. Difference of learning results under several conditions. Apexes of red pentagons represent target vowels of infant, in other words, clearest vowels in its vowel region, and black dots represent infant vowel prototypes after learning.

VI. Discussion

A. Effect of mother's sensorimotor magnets

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

To investigate the relation between the mother's sensorimotor magnets and voice cluster formation, we further simulated the interaction under several conditions where the strengths of the mother's sensorimotor magnets are different. Figure 7 shows the relationship between sensorimotor magnet strength β^c (horizontal axis) and the extent of infant voice convergence (vertical axis) during 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 were, the more tightly the infant voice clusters gather and bundle. Note that the centers of these clusters are not always the clearest vowels, as shown in Fig. 6(c).

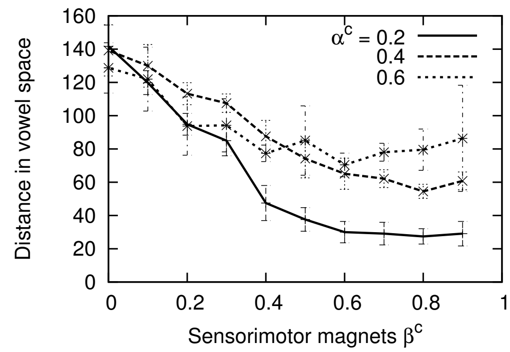


Fig. 7. Different extents of infant voice convergence during final 1000 steps in several conditions

B. Effect of mother's auto mirroring bias

We re-focus on the result shown in Fig. 6. The rating of infant vowel prototypes, which is 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 mother's auto mirroring bias helps the infant vowel prototypes approach clearer vowels, this effect probably acts with the effect of the mother'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 auto mirroring bias have the same degree.

To further investigate the effect of mother's auto mirroring bias on infant vowel prototypes, we simulated interaction in several conditions where the strengths of both auto mirroring bias and sensorimotor magnets were different. Figure 8 shows the relationship between the strengths of auto mirroring 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 density), which is the averaged value of five-time simulation with each set of α^c and β^c.

This figure indicates that the optimal strength of auto mirroring bias depends on the strength of the sensorimotor magnets. This can be explained by the features of auto mirroring bias and sensorimotor magnets. When the sensorimotor magnets are stronger, voice clusters gather and bundle more tightly at certain locations, even though the centers of these clusters are not clearest vowels. Therefore it's more probable that stronger auto mirroring bias is needed to guide such tighter voice clusters to become clear vowels.

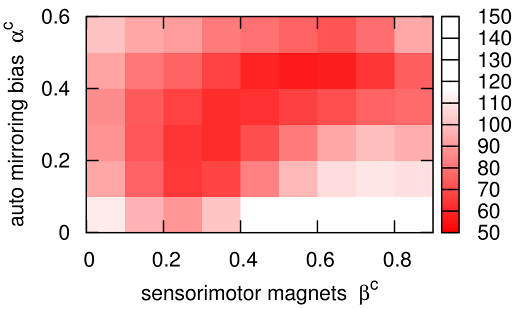


Fig. 8. Difference of ratings of infant vowel prototypes after learning in several conditions

C. Correspondence

As shown in Fig. 5, infant vowel prototypes gradually became clear vowels through mutual imitation with a mother in simulations, which follows the result of a previous experiment [3]. This result indicates the validity of considering three biasing elements, perceptual magnet effect, articulatory magnet effect, and auto mirroring bias, when constructing an imitation mechanism under a mutual imitation situation. This result also suggests that what enabled a robot to acquire clear vowels in the previous experiment was errors in human imitation derived from biasing elements. Infant studies report that real mothers and infants imitate one another's speech [1][2] and that the linguistic experience affects infant phonetic perception and vocalization[10][11][12]. Our result implies that real infants develop vowels through mutual imitation of them with their mothers, and we conclude that "errors" in mother imitations could guide infant vowel development.

VII. Conclusion

We simulated mother-infant mutual imitation of vowels with imitation mechanisms and considered two biasing elements: auto mirroring bias and sensorimotor magnets. Simulation results indicate that these biasing elements of the mother guide the infant vowel prototypes to become clear vowels; the sensorimotor magnets help form small vowel clusters and the auto mirroring bias shapes these clusters to become clearer vowels.

In simulations, we assumed that a mother always imitates an infant. However, this does not reflect real mother-infant interaction because in more realistic environment, mothers do not always perfectly imitate their infants. Therefore one of our next goals is to extend the proposed mechanism so that infants can learn vowels in conditions where caregivers do not always imitate them. Furthermore, we fixed the strength of the auto mirroring bias of the mother during interactions and assumed that a robot does not have auto mirroring bias. We will investigate how such mother and infant parameters develop. The auto mirroring bias of a mother might become stronger as the infant imitation becomes more accurate since it seems to depend on the extent to which the mother anticipates her infant imitation.

We consider that auto mirroring bias plays important roles not only in guiding infant vowel prototypes to become clear vowels but also in maintaining mother-infant interaction. We expect that auto mirroring bias forms an intra-personal positive feedback loop between the observation to be imitated and the feeling that the opponent is imitative. Yoshikawa et al. suggest the existence of inter- and intra-personal positive feedback loop not only between observation and feeling but also between feeling and action in their spiral response-cascade hypothesis [13]. They try to explain the mechanism responsible for the emergence and maintenance of communication between agents not just between a mother and an infant. We consider that auto mirroring bias is an instance of the intra-personal facilitation on interaction and mutual imitation continues with the support of the bias. We will further investigate the function of auto mirroring bias for the emergence and maintenance of mutual imitation.

References

[1] M. Peláez-Nogueras, J. L. Gewirtz, and M. M. Markham. Infant vocalizations are conditioned both by maternal imitation and motherese speech. *Infant Behavior and Development*, Vol. 19, p. 670, 1996.

[2] N. Masataka and K. Bloom. Accoustic properties that determine adult's preference for 3-month-old infant vocalization. *Infant Behavior and Development*, Vol. 17, pp. 461–464, 1994.

[3] Katsushi Miura, Minoru Asada, and Yuichiro Yoshikawa. Unconscious anchoring in maternal imitation that helps finding the correspondence of caregiver's vowel categories. *Advanced Robotics*, Vol. 21, pp. 1583–1600, 2007.

[4] P. K. Kuhl. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, Vol. 50, pp. 93–107, 1991.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B(Methodological)*, Vol. 39, pp. 1–38, 1977.

[6] Lei Xu, M. I. Jordan, and G. E. Hinton. An alternative model for mixtures of experts. In & T. K. Leen In G. Tesauro, D. S. Touretzky, editor, *Advance in Newral Information Processing System 7*, pp. 633–640. Cambridge, MA: MIT Press, 1995.

[7] Masa-aki Sato and Shin Ishii. On-line EM algorithm for the normalized gaussian network. *Neural Computation*, Vol. 12, pp. 407–432, 2000.

[8] P. K. Kuhl and A. N. Meltzoff. Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of Acoustic Society of America*, Vol. 100, pp. 2415–2438, 1996.

[9] G. E. Peterson and H. L. Barney. Control methods used in a study of the vowels. *Journal of Acoustic Society of America*, Vol. 24, pp. 175–184, 1952.

[10] P. K. Kuhl, K. A. Williams, Francisco Lacerda, K. N. Stevens, and Björn Lindblom. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, Vol. 255, pp. 606–608, 1992.

[11] K. Bloom, A. Russell, and K. Wassenberg. Turn taking affects the quality of infant vocalizations. *Journal of Child Language*, Vol. 14, pp. 211–227, 1987.

[12] K. Bloom. Quality of adult vocalizations affects the quality of infant vocalizations. *Journal of Child Language*, Vol. 15, pp. 469–480, 1988.

[13] Yuichiro Yoshikawa, Shunsuke Yamamoto, Hidenobu Sumioka, Hiroshi Ishiguro, and Minoru Asada. Spiral response-cascade hypothesis – intrapersonal responding-cascade in gaze interaction–. In *Proceedings of the Third ACM/IEEE International Conference on Human-Robot Interaction*, pp. 319–326, 2008.