# Unconscious Anchoring in Maternal Imitation that Helps Finding the Correspondence of Caregiver's Vowel Categories

Katsushi Miura*, Yuichiro Yoshikawa, Minoru Asada*

*JST ERATO Asada Synergistic Intelligence Project,*
*∗ Dept. of Adaptive Machine Systems Graduate School of Engineering, Osaka University,*
*yamadaoka 2-1 Suita-City, 565-0871,Japan*
*Katsushi.MIURA@ams.eng.osaka-u.ac.jp*
*JST ERATO Asada Synergistic Intelligence Project,*
*yamadaoka 2-1 Suita-City, 565-0871, Japan*
*yoshikawa@jeap.org*
*JST ERATO Asada Synergistic Intelligence Project,*
*∗ Dept. of Adaptive Machine Systems Graduate School of Engineering, Osaka University,*
*yamadaoka 2-1 Suita-City, 565-0871, Japan*
*asada@jeap.org*

**Abstract**

It is a formidable issue how robots can show the behaviors to be considered as the corresponding human's ones since the body structure is different between robots and humans. As a simple case for such correspondence problem, this paper presents a robot that learns to vocalize vowels through the interaction with its caregiver. Inspired by the findings of developmental psychology, we focus on the roles of maternal imitation (i.e., imitation of the robot voices by the caregiver) since it could play a role to instruct the correspondence of the sounds. Furthermore, we suppose that it causes *unconscious anchoring* in which the imitated voice by the caregiver is performed unconsciously, closely to one of his/her own vowels, and hereby helps to leading robot's utterances more vowel-like. We propose a method for the vowel learning with imitative caregiver under the assumptions in which the robot knows the desired categories of caregiver's vowels and the rough estimate of mapping between the region of sounds that the caregiver can generate and the region that the robot can do. Through the experiments with a Japanese imitative caregiver, we show that a robot succeeds in acquiring more vowel-like utterances than one without such a caregiver even when it is given different types of mapping functions.

*keywords*: maternal imitation, unconscious anchoring, dissimilar body, vowel acquisition

# 1 Introduction

It is suggested that humans generally tend to anthropomorphize the artifacts [1], and such a tendency can be amplified in facing with a humanoid robot since the similarities in appearance with humans help them to easily find the correspondences between a human and a robot. Therefore, in case of communication, humanoid robots are expected to communicate with humans in a natural manner such as vocal communication. However, it is a formidable issue how humanoid robots can show the behaviors to be considered as the corresponding humans' ones since the body structure is different from each other.

On the other hand, human infants seem to successfully solve the similar problem in the language acquisition process since infants cannot regenerate the caregiver's sounds due to sensorimotor immaturities, i.e., differences in body structure. During this process, imitation seems to have a very important role regardless of the body difference, and from the viewpoint of cognitive developmental robotics [2], the study of imitation between a human and a robot is expected not only to contribute to the studies on understanding infant cognitive development process but also to provide the robot behaviors based on these studies.

Learning to vocalize vowels seems one of the simplest tasks in imitation between dissimilar bodies since an imitator can focus only on the static features in the sound-waves to be generated. It also seems the first step of infant language acquisition that is started from only two or three months of age. Previous studies have elegantly demonstrated that a population of computer-simulated agents with a vocal tract and cochlea could self-organize shared vowels through imitations with each other [3, 4]. However, they have focused on situations that the agents have the capability to vocalize a common region instead of coping with the issue in imitation between dissimilar bodies, which is addressed in this paper.

Using a robot that can generate vowels with an artificial vocal band and tract (ex. [5, 6]) is one approach to directly attack the problem of imitation between dissimilar bodies. With such a vocal robot, Yoshikawa et al. [7] proposed a mother-infant interaction model for infant vowel acquisition based on the observations in developmental psychology. Inspired by the findings that maternal imitation effectively reinforces infant vocalization [8] and that its speech-like cooing tends to invoke utterances by its mother [9], they have suggested that maternal imitation (i.e., imitation of the robot's utterance by the caregiver) using adult phonemes plays an important role in phoneme acquisition, namely matching its articulations and the corresponding caregiver's utterances. In their model, the robot could find a lot of candidates of vowels and could determine which of them are more vowel-like, in other words which of them are easier for humans to recognize as vowels.

We suppose that the maternal imitation could play another important role in vowel learning beyond the role of giving the instruction to match the caregiver's vowel to the robot's utterance, that is leading the robot's utterance to be more vowel-like. In this paper, we present an interaction paradigm and experiments in order to show this another role of the maternal imitation.

The test task for a vocal robot is learning how to articulate vowel-like sounds through the interaction with a caregiver who tries to imitate the robot utterances but cannot regenerate them due to the difference between their articulatory systems. In this setup, it is conjectured that the imitated voice by the caregiver is performed unconsciously, closely to one of his/her own vowels, and we call such a behavior ”unconscious anchoring”. Maternal imitation and this unconscious anchoring would cause two phenomena that support learning of more vowel-like sounds: (1) given maternal imitation, the robot can obtain the references to modify the mapping between the sound feature vectors of vowels generated by the caregiver and that by the robot, and (2) furthermore, by unconscious anchoring, the references would be gradually shifted to more vowel-like sounds.

In the rest of this paper, we introduce the idea of unconscious anchoring and a learning mechanism based on it. Through some experimental trials of vowel learning with a Japanese imitative caregiver, we show that the robot succeed in acquiring more vowel-like utterances compared to the robot utterances without such a caregiver. Furthermore, we show that the robot could learn vowel-like sounds even when the robot is given different types of mapping functions.

## 2 Assumptions and basic ideas in unconscious anchoring

An interaction model between a caregiver and a robot is shown in Figure 1 where a vocal robot interacts with a caregiver through vocalization and hearing the caregiver’s voices. In this scenario,

**R:** the robot tries to utter one of Japanese vowels, and

**C:** the caregiver listens to the robot’s utterance, looks at the shape of robot lip, and then tries to imitate the voice of the robot.

Such imitation by a caregiver is expected to provide the robot with information on how the voice of the robot is interpreted by the caregiver, which seems to reveal the most important aspects with regard to achieving communication.

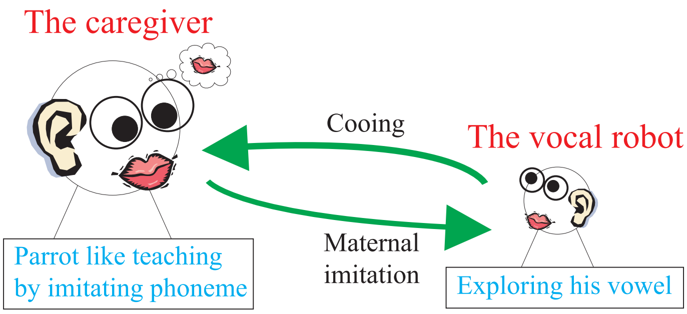


Figure 1: An interaction model between an imitative caregiver and a vocal robot

The task of the robot through such interaction is learning to find ways of articulation by which it can generate sounds corresponding to the caregiver’s vowels. The robot cannot generate the exactly same sound as the caregiver's one and vice versa since the articulatory system is different from each other. In other words, the regions of sounds that the caregiver and the robot can generate are usually different from each other or, at worst, do not overlap with each other at all. Nevertheless, humans can map the robot's sounds to their own corresponding vowels [7]. In contrast, it is usually not trivial for the designers to provide their robots with the accurate mappings between these two regions of sounds. Therefore, we assume that we can provide only the rough estimates for the mapping function.

The human's utterances can be clustered in the space of the static feature of sound-wave, namely *formants*, in which clusters correspond to vowels [10]. Therefore, it is reasonable to assume that we can provide the robot with the categories of the desired vowels or that the robot learns them through the observation of human's usual utterances.

Given the above assumptions, when it listens to the caregiver's imitative utterance, the robot can obtain information how its attempting voice differs from the desired vowel category. Then, it can obtain rough information of the difference in its own regions of sounds by using the mapping function. The mapped difference can be used for modifying its own 'vowel' category. The phenomenon of leading the robot's utterance to be more vowel-like would occur by virtue of the following implicit assumption underlying the mutual imitation process. While he/she attempts to imitate the robot's voice, the caregiver unconsciously uses his/her own voice and vowel due to the sensorimotor constraints. In other words, the caregiver's imitative voice is slightly biased towards the direction of his/her own vowel category. Consequently, since the directions of modifying the robot's categories are biased towards ones corresponding to the caregiver's vowels, the robot sounds would gradually become more vowel-like, i.e., easier for humans to recognize them as vowels.

The idea of unconscious anchoring would be generalized to other modalities such as vision (and motion that is, gesture) and hopefully guide a new methodology of providing a robot with social skills through interaction. As a first step, in this paper, we focus on the issue of robot vocal acquisition through mutual imitative interaction with vision and audition.

## 3 Learning method

The robot learns how to articulate the vowels corresponding to the caregiver's ones through mutual imitation. In the learning process, the 'vowel' categories of the robot defined in the *formant space* are updated through the interaction with an imitative caregiver. In this section, we introduce how we provide the robot with rough estimation of the mapping by which it can convert the information of the correspondence onto the region of its own generable sound. We then introduce the updating rule of the 'vowel' categories of the robot.

### 3.1 Mapping functions between the regions of generable sounds

Human's vowels are well distinguished in the formant space, a well-known sound feature space for vowel classification [10]. Figure 2 shows sample distributions of five Japanese vowels uttered by a Japanese male and a Japanese female. As you can see from Figure 2, the categories of Japanese vowels are distributed in the formant space as if they form a pentagon.

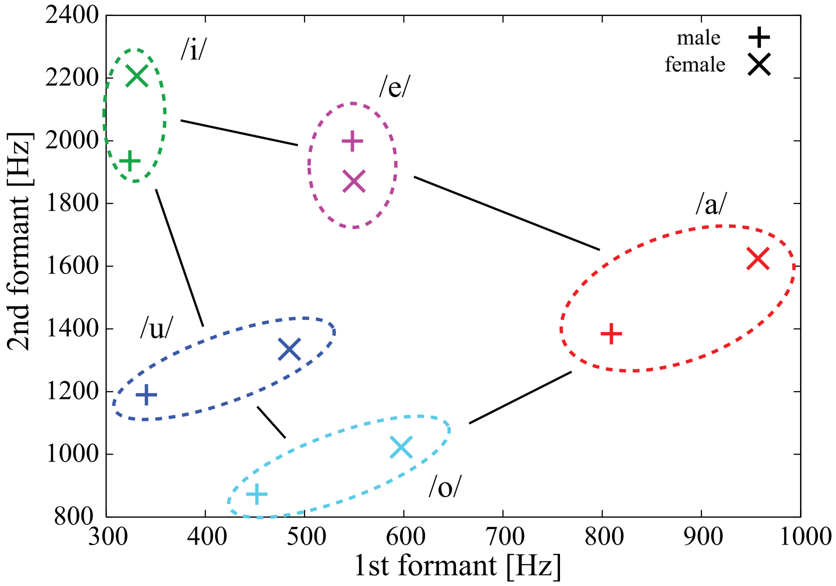


Figure 2: A sample distribution of human vowels in the formant space

Since we suppose that forming a pentagon in the formant space is an important feature for vowel categories, possible pentagons in the regions of generable sounds by the robot are expected to be feasible starting positions for learning. Therefore, we provide the robot with a linear transformation as a mapping function from the region of generable sounds by the caregiver to one by the robot. In other words, a caregiver's vowel **h** is converted to the sound corresponding **h**' by a mapping function **g** with the parameters of a scaling coefficient $\alpha$, a rotational matrix $\mathbf{R}(\theta)$ by the angle $\theta$, and an offset vector **s** such as

$$\mathbf{h}' = \mathbf{g}(\mathbf{h}; \alpha, \theta, \mathbf{s}) \equiv \mathbf{r}_c + \alpha \mathbf{R}(\theta)(\mathbf{h} - \mathbf{h}_c) + \mathbf{s} \tag{1}$$

where $\mathbf{h}_c$ and $\mathbf{r}_c$ indicate the centroids of the generable regions by the caregiver and the robot, respectively.

### 3.2 Updating the 'vowel' categories of the robot

The imitated voice of the robot utterance by the caregiver is supposed to tell the difference of the robot utterance from the sound of the closest vowel category of the caregiver. The differences can be converted to the ones by the robot based on the mapping function and be used to update the 'vowel' categories of the robot.

Suppose that the robot utters $\mathbf{r}_d^{/v/}$ $(/v/ = /a/, /i/, /u/, /e/,$ or $/o/)$ that is one of the current prototype vowel category of $/v/$ and the caregiver generates the imitated sound $\mathbf{h}$. Let the prototype category of the usual caregiver's vowel $/v/$ be $\mathbf{h}^{/v/}$. The robot updates $\mathbf{r}_d^{/v/}$ based on the difference between $\mathbf{h}$ and $\mathbf{h}^{/v/}$. These processed are formalized as the following way:

1. At the $k$-th step, the robot selects one of vowels $/v/$ and utters it with the current prototype category $\mathbf{r}_d^{/v/}(\mathrm{k})$.

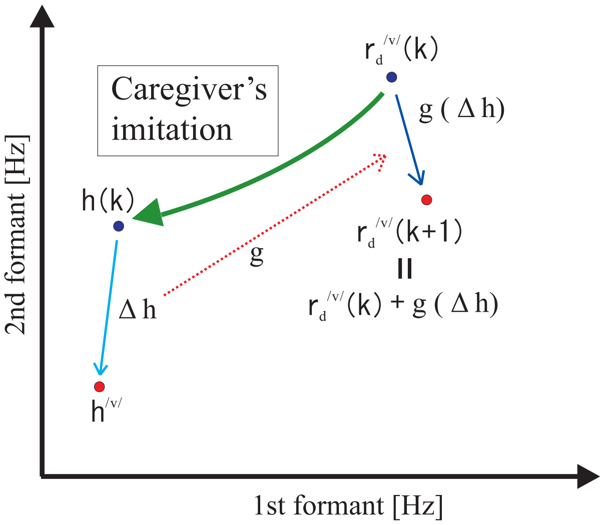
2. The caregiver generates the sound $\mathbf{h}(\mathrm{k})$ to imitate the robot's utterances.

3. The difference vector $\Delta\mathbf{h} = \mathbf{h}^{/v/}$ - $\mathbf{h}(\mathrm{k})$ is converted to the region of generable sounds by the robot with the mapping function $\mathbf{g}$. The converted difference vector is applied to modify the prototype category $\mathbf{r}_d^{/v/}(\mathrm{k})$, in other words,

$$\mathbf{r}_d^{/v/}(k+1) = \mathbf{r}_d^{/v/}(k) + \mathbf{g}(\Delta\mathbf{h}) \tag{2}$$

   Figure 3 illustrates these processes in a schematic way.

4. Again, the robot utters the voice with the new prototype category $\mathbf{r}^{/v/}(k+1)$.



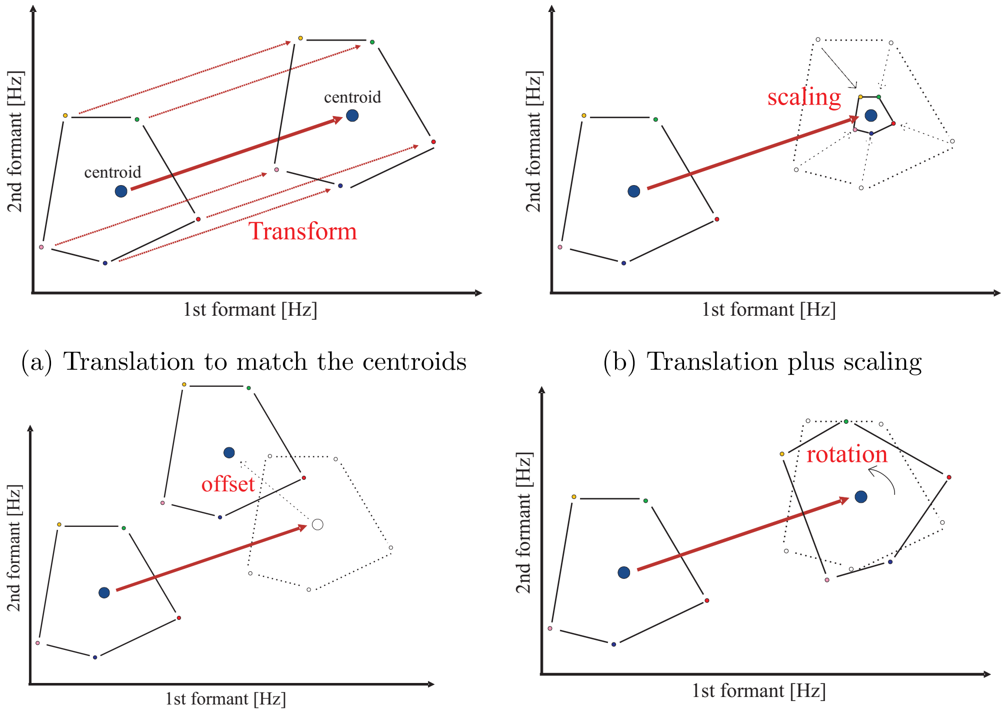
$\mathrm{r}_\mathrm{d}^{/v/}(\mathrm{k})$
Caregiver's imitation
g ( Δ h)
h(k)
g
$\mathrm{r}_\mathrm{d}^{/v/}(\mathrm{k+1})$
II
$\mathrm{r}_\mathrm{d}^{/v/}(\mathrm{k}) + \mathrm{g}(\Delta\mathrm{h})$
Δ h
$\mathrm{h}^{/v/}$
2nd formant [Hz]
1st formant [Hz]

Figure 3: Updating process of the prototype vector of a vowel $(/v/)$ category $\mathbf{r}_d^{/v/}$

# 4 Experiment

In the experiments, we like to verify our hypotheses on the role of maternal imitation in the acquisition process of more vowel-like utterances by the robot: (1) the imitated sounds by the caregiver converge on his/her own vowels owing to "unconscious anchoring" regardless of different mapping functions, (2) the vowels that the robot acquired through the maternal imitation are more acceptable as Japanese vowels than ones acquired from the fixed desired formant vectors.

We used four types of rough estimation in the experiments as shown in Figure 4: (a) translation to match the centroids, (b) translation plus scaling, (c) translation plus offset, and (d) translation plus rotation.



(a) Translation to match the centroids

(b) Translation plus scaling

(c) Translation plus offset

(d) Translation plus rotation

Figure 4: Four examples of the mapping functions from the region of generable sounds by the caregiver to one by the robot

First, we present our vocal robot and explain how it utters. Next, the experimental procedures are explained, and then the results on the imitated and acquired vowels with statistical analysis are given.

## 4.1 The vocal robot

Vocalization is commonly regarded as the result from a modulation of a source of sound energy by a filter function determined by the shape of the vocal tract; this is often referred to "source-filter theory of speech production" [11] and implemented also in the previous studies [7, 6]. To model the process of vowel convergence in mother-infant interaction, we improved the vocal robot used in previous study [7] in such a way that we replaced the sound source with an air compressor and an artificial vocal band, and added a lip at the front end of the vocal tract, and the length of the robot's vocal tract changes from 170 [mm] (male's average vocal tract length) to 116 [mm].

Figure 5, 6 shows the new vocal robot and articulatory system of it. The compressed air is conveyed through a tube to the artificial vocal band to generate the source sound of fundamental frequency (see Figure 7); then the sound-wave is spread out through the vocal tract and the lip, that is a silicon tube with hollow end which resembles a human lip (see Figure 8). To modulate the sound-wave, the vocal tract and lip were wired with four electric motors, respectively, and could be deformed by them. The host computer controls the motors through motor controllers (usbMC01, iXs Research Corp.). The host computer receives signals from a microphone and calculates their formants.

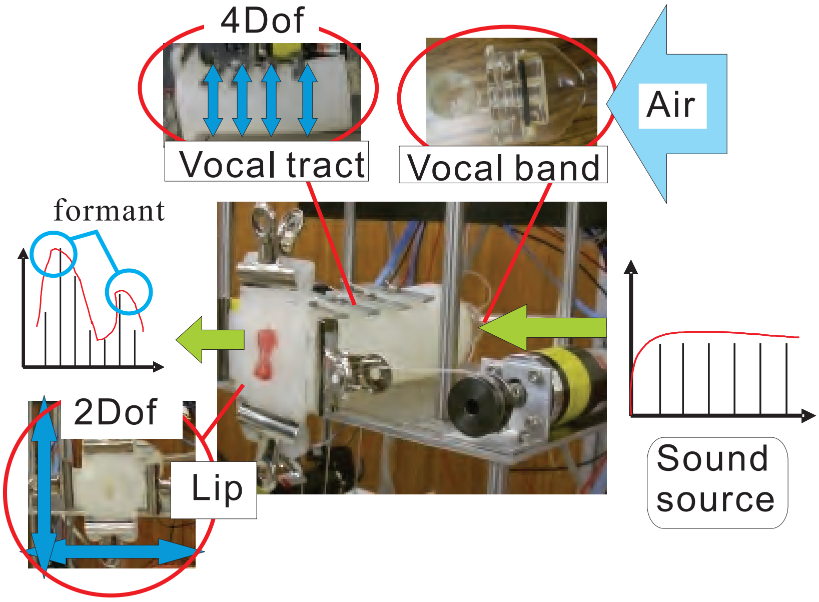


Figure 5: The articulatory system of the vocal robot

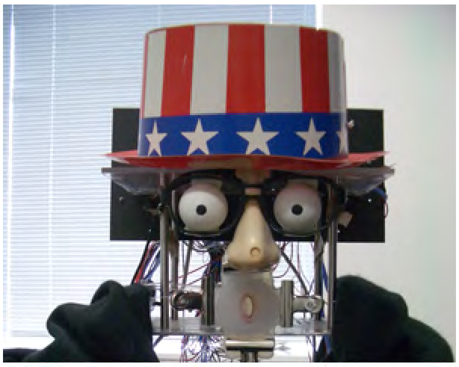


Figure 6: The vocal robot

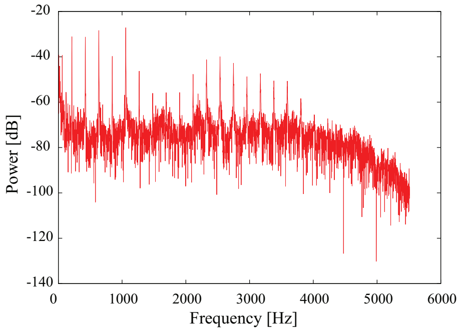
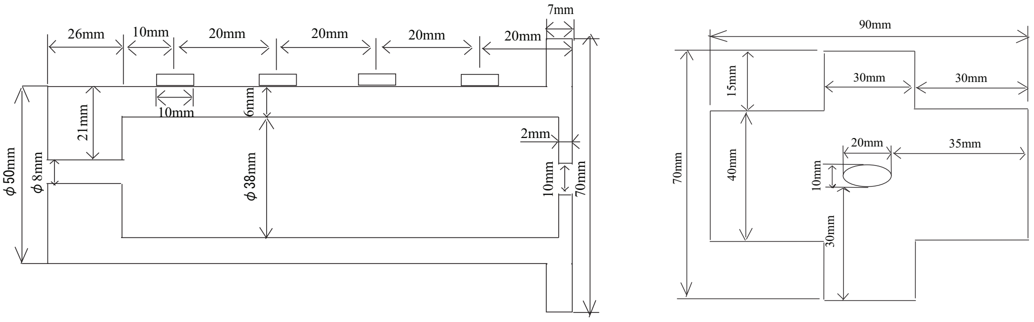


Figure 7: The spectra of the sound source generated by the artificial vocal band

The vocal robot has six degrees of freedom, two of which are used for opening/closing of lips by four motors, and four of which for deforming the vocal tract by another set of motors. First, we present the utterance capability of the robot. The motor commands which control the shape of vocal tract are quantized into five levels, 0 (free, no deformation), 0.25, 0.5 (medium), 0.75, and 1.0 (maximum deformation), and the motor commands which control lip shape are assigned to imitate the shape of human lips. Table 1 shows the motor commands used to imitate human lip shape, and Figure 9 shows the formant distribution of the robot utterances in the formant space where the horizontal and vertical axes indicate the first and second formants, respectively. Furthermore, Figures 10 (a), ⋯, and (e) indicate



(a) Side view

(b) Front view

Figure 8: Sizes of the vocal tract

the formant distributions categorized by the lip shape. In Figure 10 formant distribution relates with the robot's lip shape, and the larger opening size is, the higher the first and second formant shift.

Table 1: The motor commands to form the lip shapes to resemble human's ones in vocalizing vowels

| Motor output to control the robot's lip shape | /a/ | /i/ | /u/ | /e/ | /o/ |
|---|---|---|---|---|---|
| vertical direction | 1.0 | 0.0 | 0.0 | 0.5 | 0.5 |
| horizontal direction | 1.0 | 1.0 | 0.0 | 1.0 | 0.0 |

By using the data in Figure 9 as the list of pairs of motor commands and formant vectors, the robot can generate the desired sound. From the list of the pairs, it can finds some candidate pairs of which formant vectors are sufficiently close to the desired one. Then, it selects a pair from the candidates, which has the closest motor command to the previous motor one.

## 4.2 Set up and procedure

The experiments are conducted on the condition where one subject (the same caregiver through the all experiments) participated in the vowel acquisition process by two kinds of methods with four kinds of mapping functions each, that is, totally eight kinds of experiments to verify the hypotheses. Note that each experiment is iterated five times for the later statistical analysis. In the vowel acquisition process with the proposed method of maternal imitation, the caregiver tries to imitate the robot's utterances as other person would judge whether his imitated utterance was the same as the robot's one. Through turn taking of uttering voice, the robot modifies the desired formant vectors by using the caregiver's utterances as an indication of the correspondence of both utterances. For comparison, another process of vowel acquisition is performed by a supervised learning method with fixed desired formant vectors specified by the mapping function. The variations of the four kinds of mapping functions are as follows:

**(translation)** only translation by the difference between two centroids: $\alpha$=1.0, $\mathbf{R}(0)$, $\mathbf{s}$=(0, 0) (See Figure 4 (a)).

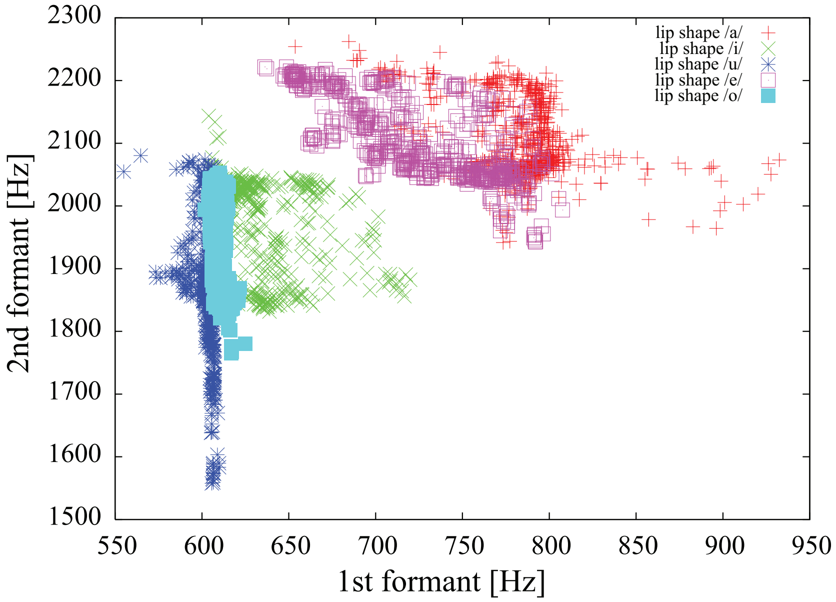


Figure 9: The distribution of the robot utterances in the formant space

**(scaling)** translation to match the centroids plus scaling: $\alpha$=0.24 (that coincides with the region of the generable sounds of the robot), $\mathbf{R}(0)$, $\mathbf{s}$=(0, 0) (See Figure 4 (b)).

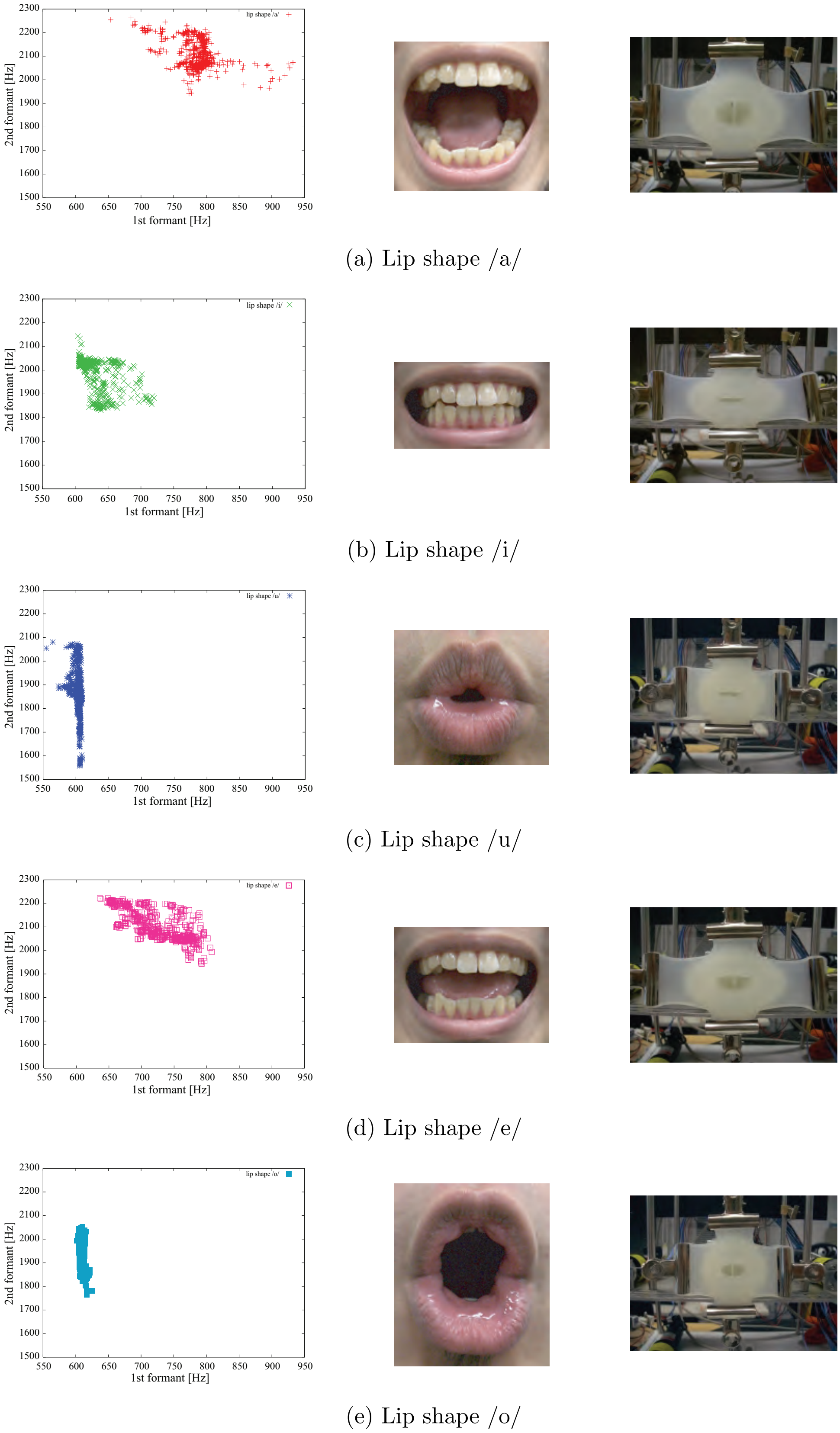
**(offset)** translation to match the centroids plus offset: $\alpha$=1.0, $\mathbf{R}(0)$, $\mathbf{s}$=(-100, 200) (See Figure 4 (c)).

**(rotation)** translation to match the centroids plus rotation: $\alpha$=1.0, $\mathbf{R}(30)$, $\mathbf{s}$=(0, 0) (See Figure 4 (d)).

The mapping function of 'translation' is regarded as one of simple and feasible translation to match vowels in different regions of formant space. The other mappings are example varieties which contain some noise in such a feasible mapping. Note that the number of steps for the supervised learning and the number of turn takings for the maternal imitation are 20 for each vowel category.

### 4.3 Results

First, we present the robot's vowels that acquired through experiments. Figure 11 shows the vowel categories in the formant space that the robot acquired with the supervised learning and the maternal imitation with a mapping function (translation). In Figure 11 (a), the desired formant vectors in the supervised learning and the final desired formant vectors modified in the proposed learning process with the maternal imitation are indicated as blue symbols (+, * etc.) and red ones. Hereafter, blue and red colors indicate the data by the supervised learning and the maternal imitation, respectively. In Figure 11 (b), the vowel categories as formant vectors acquired by the both methods are indicated in the same colors as Figure 11 (a). Figures 12 (a) and (b) show the similar graphs as Figures 10 (a) and (b) in the case of averaging the vowel categories among the four mapping functions. The differences between the supervised learning and the learning with maternal imitation in Figure 10 and Figure 11 imply that the



(a) Lip shape /a/

(b) Lip shape /i/

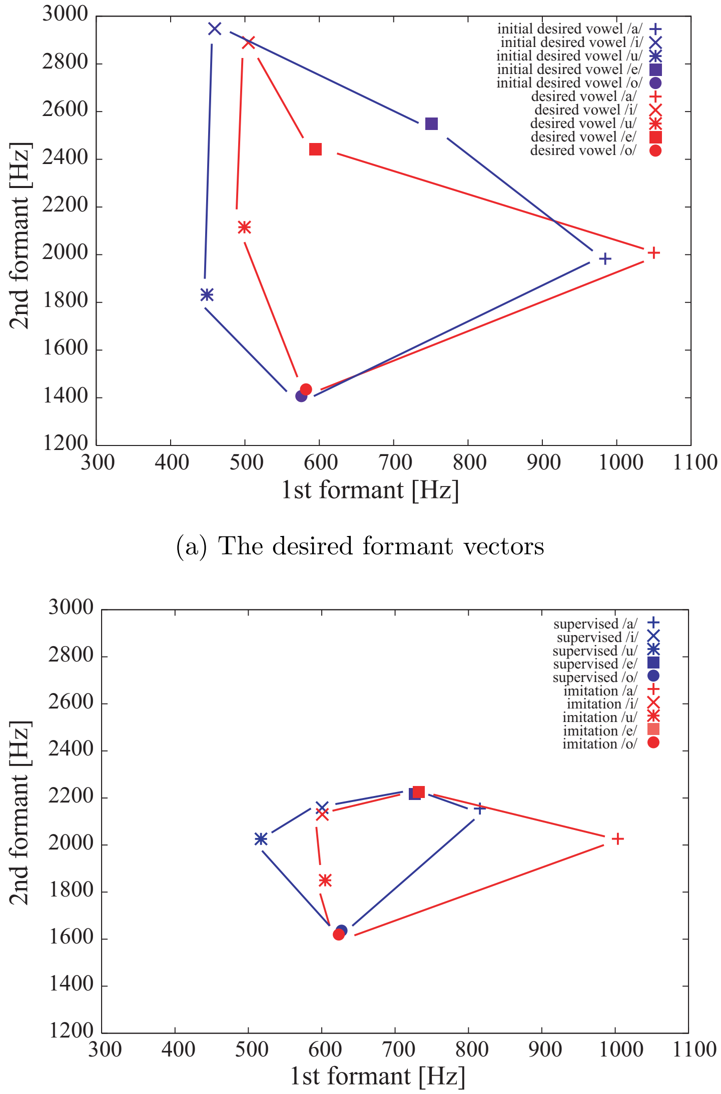
(c) Lip shape /u/

(d) Lip shape /e/

(e) Lip shape /o/

Figure 10: The distributions of the robot utterances in the formant space each of which is generated with a lip shape that resembles human's one for the corresponding vowel utterance

robot succeeded in modifying its desired formant vectors.

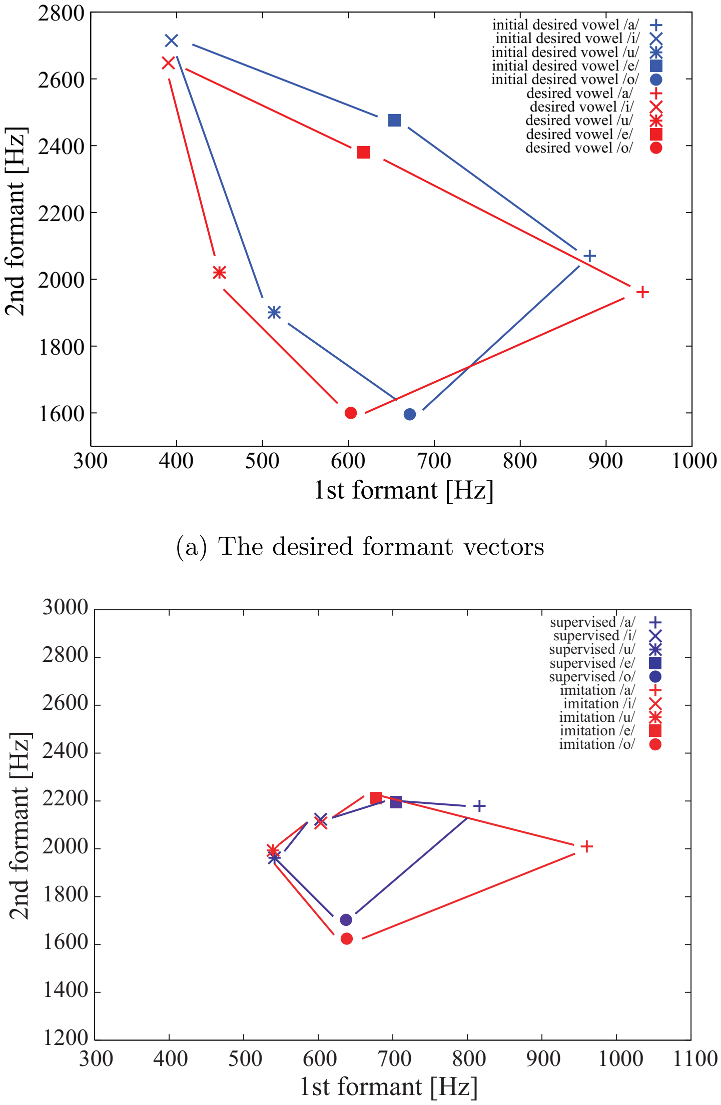


(a) The desired formant vectors

(b) The acquired formant vectors

Figure 11: The vowel categories in the formant space that the robot acquired by the supervised learning and the maternal imitation with a mapping function (translation)

We hypothesized that the unconscious anchoring gradually leads the caregiver's utterance to his/her own vowels, and to check this tendency, the changes of the difference $\Delta\mathbf{h}$ in Figure 3 (the distance indicating the error of the mapping) at the beginning and at the end of the learning are examined. This change is shown in Figure 13 where the vertical axis indicates the size of $\Delta\mathbf{h}$ and the vertical bars indicate average of first three times learning of ones and average of last three times learning, respectively. Each bar was acquired through five times experiment with each mapping functions. The narrow bars indicate standard deviation of them. From the T-test, there appeared to be highly significant difference of the average size of $\Delta\mathbf{h}$ between in the first three steps and in the last three steps ($p = 2.0 \times 10^{-5}$).



(a) The desired formant vectors

(b) The acquired formant vectors

Figure 12: The average vowel categories acquired by the supervised learning and the maternal imitation

This difference implies the tendency of the convergence of the caregiver's imitation, hopefully to his/her own vowels. This result seems to support the verification of the first hypothesis.

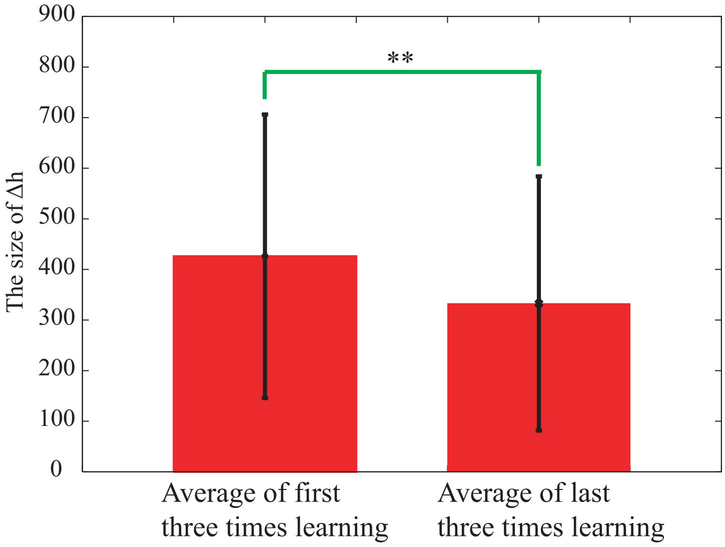


Figure 13: The difference between the imitated voices by the caregiver and his/her corresponding usual vowels at the beginning and at the end of the interaction : ** indicates the highly significant difference between them

Since it is difficult to show which is more vowel-like between two methods (the maternal imitation and the supervised learning) in the formant space, we utilized a subjective criterion to judge it. We made 15 subjects compare the vowels acquired through maternal imitation with the vowels acquired by the supervised learning. In this test, the robot continuously utters vowels acquired through maternal imitation or with the supervised learning with four different mapping functions in the normal order of Japanese vowels, that is /a/,/i/,/u/,/e/,/o/. Subjects compare the robot's vowels four times (two sets of the voices by the maternal imitation followed by that of the supervised learning and vise versa) in terms of four kinds of mapping functions and judge which were more like Japanese vowels.

Figure 14 shows the result of the comparison by 15 subjects where the percentage of subjects who answer the vowels acquired by the maternal imitation is better with each mapping function and the total percentage among all four mapping functions are indicated as red bars and the percentage of subjects who answer the vowels acquired with supervised learning is better are indicated as blue bars. We conduct the tests whether the percentage of subjects who positively answer on the utterances acquired by the maternal imitation is larger than the chance level (50%). From the statistical tests, we found that the subjects tend to judge the utterances acquired by the maternal imitation more vowel-like than the ones acquired by the supervised learning for three mapping functions, namely translation ($p = 1.6 \times 10^{-3}$), offset ($p = 4.7 \times 10^{-5}$), and rotation ($p = 2.2 \times 10^{-2}$). Although there was not the significant difference with the chance level in the case of a mapping function of scaling ($p = 2.8 \times 10^{-1}$), the test on total percentage among four mapping functions indicated that the tendency of the utterances acquired by the maternal imitation as more vowel-like sound ($p = 1.1 \times 10^{-2}$). From the result of the comparison, we may conclude that the robot could acquire vowels with the maternal imitation, which are better, or at worst equally well to recognize as Japanese vowels for human regardless of the mapping functions.

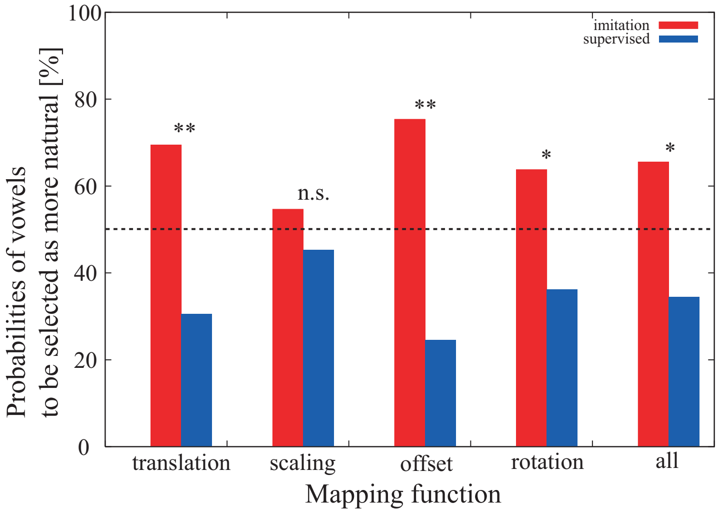


Figure 14: The probabilities of vowels to be selected as more natural: the red, left-side bars indicates the probabilities of ones acquired by the proposed imitative learning while the blue, right-side bars indicates ones acquired by the supervised learning with four types of rough estimation of the mapping, * indicates the significant difference between the percentage of subjects who positively answer on the utterances acquired by the maternal imitation and chnace level, ** indicates the highly significant difference, n.s. indicates no significant difference.

## 5 Discussion and conclusion

In this paper, we propose a vowel learning method with an imitative caregiver based on the idea of unconscious anchoring behavior of humans. In the experiment, a vocal robot could succeeded in acquiring more vowel-like sounds through the iterative turn-taking with an caregiver who tries to imitate the robot's utterances. Even though the caregiver did not intend to instruct the robot but did intend to just imitate the robot's sound, the human tendency of unconscious anchoring worked to lead the robot's utterances to more vowel-like. Since utilizing unconscious anchoring is expected to be a new paradigm to provide a robot with a human-like behaviors, we should quantitatively investigate to what extent such a tendency can be expected and can be used for this purpose.

One of the most fundamental issues in imitation is to find the mapping function from the observation of other's behaviors to own behaviors. Here the mapping function is approximated with an affine one in the formant space. The parameters used in the experiments are limited and we have not tried other parameters. Too much wrong estimates of the affine approximation does not seem work, but we suppose that the parameters might work unless the vowel categories are interfered each other by the transformation for modification. If they interfered with each other, the desired formant vector would jump to the wrong category. How can we guarantee no interfere is one of our future issues.

In our study, the parameters of the mapping function does not change during the maternal imitation process although these parameters are not exactly correct. Therefore, learning (modification) of these parameters simultaneously with modification of the desired vectors is a natural extension of the current work, and we conjecture that it is what an infant does during his/her cooing process knowing the vowel categories from the experiences of listening to his/her mother's utterances. The supporting facts are partially observed in developmental psychology [12, 13]. However, real infants are exposed by not simply single vowels but rather continuous voices with consonants, that is, words, phrase, and sentences. Further, mothers do not simply respond by imitation of their infants' utterances. In such an environment, how can our current work be extended is a very challenging issue.

Since "unconscious anchoring" is considered as the general concept of human behaviors in imitation, the framework of the current work is expected to be applied to other modality such as vision (and motion, that is, gesture). To show this generality is another our future issue. Since the multi-modal sense of human is said to be interfered with each other (such as McGurk effect [14]), another possibility of the extension would be concerning the hybrid effects of unconscious anchoring not only in auditory but also in vision. Although the effects of these modalities were not well separated in the current work, investigations to separate these effects would be an important issue and hopefully derive some requirements of the body or appearance of the robot to effectively utilize the multi-modal unconscious anchoring.

# REFERENCES

[1] Byron Reeves and Clifford Nass. *The media equation -how people treat computers, television, and new media like real people and places.* Stanford Univ Center for the Study, 1996.

[2] Minoru Asada, Karl F. MacDorman, Hiroshi Ishiguro, and Yasuo Kuniyoshi. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous System*, 37:185–193, 2001.

[3] B. de Boer. Self-organization in vowel systems. *Journal of Phonetics*, 28:441–465, 2000.

[4] P.-Y. Oudeyer. Phonemic coding might result from sensory-motor coupling dynamics. In *Proceedings of the 7th international conference on simulation of adaptive behavior (SAB02)*, pages 406–416, Edinburgh, UK, August 2002.

[5] Kotaro Fukui, Kazufumi Nishikawa, Shunsuke Ikeo, Eiji Shintaku, Kentaro Takada, Hideaki Takanobu, Masaaki Honda, and Atsuo Takanishi. Development of a talking robot with vocal cords and lips having human-like biological structurest. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2526–2531, Augst 2005.

[6] T. Higashimoto and H. Sawada. Speech production by a mechanical model construction of a vocal tract and its control by neural network. In *Proc. of the 2002 IEEE Intl. Conf. on Robotics & Automation*, pages 3858–3863, Washington D.C., U.S.A, May 2002.

[7] Yuichiro Yoshikawa, Minoru Asada, Koh Hosoda, and Junpei Koga. A constructivist approach to infants' vowel acquisition through mother-infant interaction. *Connection Science*, 15(4):245–258, Dec 2003.

[8] M. Peláez-Nogueras, J. L. Gewirtz, and M. M. Markham. Infant vocalizations are confitioned both by maternal imitation and motherese speech. *Infant behavior and development*, 19:670, 1996.

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

[10] R. K. Potter and J. C. Steinberg. Toward the specification of speech. *Journal of the Acoustical Society of America*, 22:807–820, 1950.

[11] Philip Rubin and Eric Vatikiotis-Bateson. *Animal Acoustic Communication*, chapter 8 Measuring and modeling speech production. Springer-Verlag, 1998.

[12] Anthony J. DeCasper and Melanie J. Spence. Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, 9:133–150, 1986.

[13] Patricia K. Kuhl. Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, 66:1668–1679, 1979.

[14] Harry McGurk and Jrohn MacDonald. Hearing lips and seeing voices. *Nature*, 72:746–748, 1976.