

Towards Emergence of Communication: A Cognitive Developmental Robotics Approach

Minoru Asada

JST ERATO Asada Synergistic Intelligence Project and Dept. of Adaptive Machine Systems
Graduate School of Eng., Osaka University, Suita, Osaka 565-0871, Japan

asada@jeap.org

ABSTRACT

This paper presents an introduction of the JST ERATO Asada Synergistic Intelligence Project (hereafter, SI) and its preliminary studies on emergence of communication. The project aims at building cognitive developmental artificial agents (humanoids), understanding natural agents (humans), and their mutual feedbacks. The project consists of four groups: (1) Physio-SI: whole body dynamic motions such as walking, running, and jumping, (2) Perso-SI: cognitive developmental robotics including body image, imitation, and language communication, (3) Socio-SI: emergence of communication and society by androids, and (4) SI-mechanism: neuroscientific supports for Physio, Perso, and Socio-SIs. As a preliminary study, vowel acquisition based on visual and auditory mutual imitation is given. It aims at modelling mother-infant interaction in vowel acquisition based on the findings in infant studies, such as lip shape imitation and bias from mother's vocalization onto the infant auditory perception.

1. INTRODUCTION

The recent advances in information and robot technology (in short, IRT) enable us to build humanoid robots that have a large number of DoFs and various kinds of many sensors. Human-like motions are realized on these robots, and the shape and appearance have become closer and closer to us. [12, 4].

However, the current robotics lacks the faculties of language communication with ordinary people and of intelligent behavior generation in various situations such as at home. The fundamental relationship between humans and robots would become more important since these robots would be introduced into our lives in near future, and therefore the mechanisms of adaptation and development of both humans and robots should be taken into account in order to find the correct direction of the technology in future.

"Synergistic Intelligence (hereafter, SI)," the title of JST (Japan Science and Technology Agency) ERATO (Exploratory

Research for Advanced Technology) Asada project, emerges intelligent behaviours through the interaction with environment including humans. Synergistic effects with brain science, neuroscience, cognitive science, and developmental psychology are expected. SI is one approach to a new discipline called "Humanoid Science" that aims at providing a new way of understanding ourselves and a new design theory of humanoids through mutual feedback between the design of human-like robots and human-related science.

"Humanoid Science" under which a variety of researchers from robotics, AI, brain science, cognitive science, psychology and so on are seeking for new understanding of ourselves by constructivist approaches, that is expected to produce many applications.

SI adopts a methodology called "Cognitive Developmental Robotics" (hereafter, CDR) [1] that consists of the design of self-developing structures inside the robot's brain, and the environmental design: how to set up the environment so that the robots embedded therein can gradually adapt themselves to more complex tasks in more dynamic situations. Unstructured terrains are opponents for adaptive walkers to negotiate with in order to generate dynamic motions. The caregiver's behaviour to a robot is one environmental design issue since parents, teachers, and other adults adapt themselves to the needs of children according to each child's level of maturity and the particular relationship they have developed with that child.

One of the most formidable issues in SI is "Nature vs. Nurture": to what extent should we embed the structure, and to what extent should we expect the development triggered by the environment? A symbolic issue is "Language Acquisition." "How can robots emerge the symbol in the social context? What is the essential element in this process?"

This paper presents an introduction of SI and its preliminary studies on emergence of communication. The project aims at building cognitive developmental artificial agents (humanoids), understanding natural agents (humans), and their mutual feedbacks (see Fig.1). The project consists of four groups: (1) Physio-SI: whole body dynamic motions such as walking, running, and jumping, (2) Perso-SI: cognitive developmental robotics including body image, imitation, and language communication, (3) Socio-SI: emergence of communication and society by androids, and (4) SI-mechanism: neuroscientific supports for Physio, Perso, and Socio-SIs. As a preliminary study, vowel acquisition based on visual and auditory mutual imitation is given. It aims at modelling mother-infant interaction in vowel acquisition based on the findings in infant studies, such as lip

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AAMAS'06 May 8–12 2006, Hakodate, Hokkaido, Japan.
Copyright 2006 ACM 1-59593-303-4/06/0005 ...\$5.00.

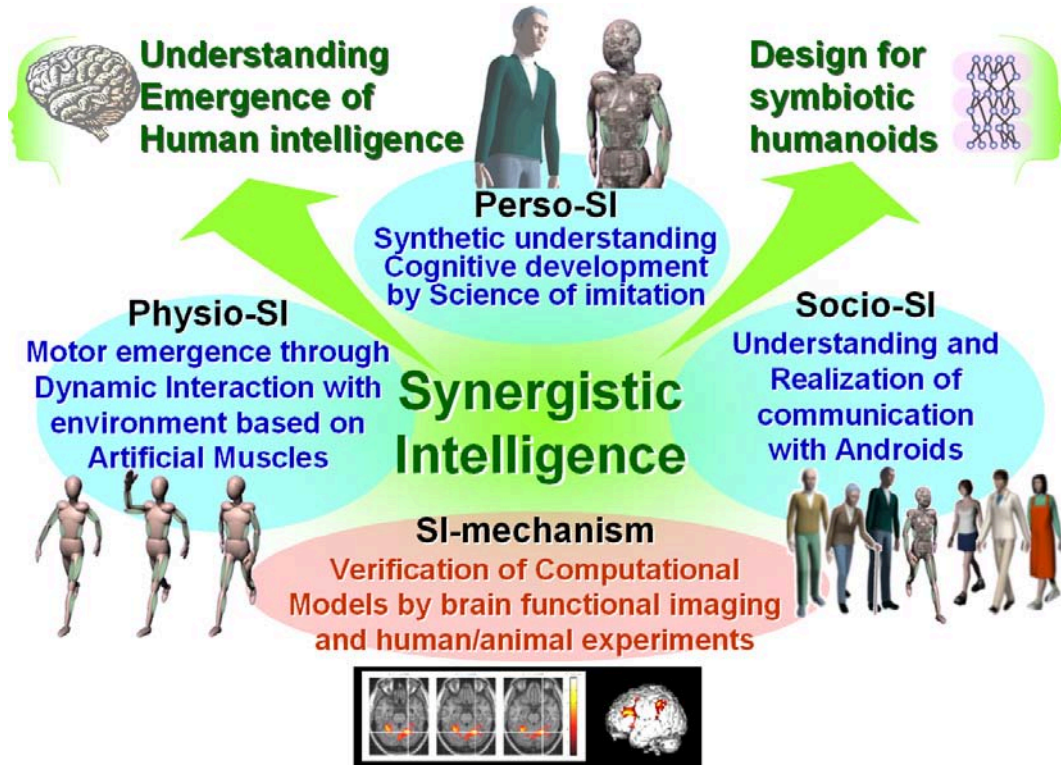


Figure 1: An overview of Synergistic Intelligence Project

shape imitation and bias from mother’s vocalization onto the infant auditory perception.

2. AN OVERVIEW OF SI

In this project, we will establish a new discipline that explains the essential relationship between humans and robots in future symbiotic society. Systematic and continuous modelling of “embodiment,” “autonomy,” and “sociality” from a viewpoint of emergence and development will be performed, and constructivism will be shown by realizing these models on real robots. Fundamental principles for adaptive and developable robots will be revealed and may contribute to human society.

An expected goal is realization of three years old humanoid that can show the capabilities of 1) seamless transition among running, jumping, walking, and stop, and whole body motion such as rope skipping, 2) naming game with toy playing and novel behavior generation from imitation, and 3) self-expression, attention, and conversation. The rest of this section gives overviews of four groups of SI.

2.1 Physically Synergistic Intelligence

Physio-SI in short, aims at understanding emergence of human intelligence from dynamic connection between body structure and environment with artificial muscles.

The agent behavior is determined by not only its dynamics but also environmental one if its compliance is comparable to that of the environment. This interaction may help the agent to be adaptive and robust against changes of the environment, which cannot be taken for the rigid agents with high-ratio geared motors. In this project, we utilize

such physical interactions to realize dynamic skills such as walking, jumping, and running, and try to figure out the methodology to emerge synergistic intelligence:

- **Walking on uneven terrains utilizing antagonistic driven legs:** By utilizing flexibility provided by antagonistic driven legs and ballistics of the body, we will derive a methodology to realize adaptive walking on uneven terrains Preliminary results are shown in [11].
- **Realization of dynamic behaviors:** We will realize dynamic behaviors such as jumping and running by utilizing compliance and dynamic power provided by the antagonistic drive.
- **Smooth transition between dynamic behaviors:** We will propose a framework to deal with different modes of locomotion, walking and running, and develop a “Motor Skill Emerging Humanoid” that can realize these locomotion modes.
- **Cognitive approach to locomotion:** From the standpoint of building a Motor Skill Emerging Humanoid, we approach to the synergistic intelligence mechanisms and will understand the foundation of emergence of dynamic behaviors.

2.2 Interpersonally-Synergistic Intelligence

Fusing together the capabilities of real world behavior and communication will be the critical step towards the truly humanoid robot intelligence. Interpersonally-Synergistic Intelligence group (Perso-SI in short) aims at constructing

synthetic robotic models that account for the process of cognitive development starting from bodily movements of fetuses and reaching rich meaningful communication of infants. This will bridge the research of the other two groups, namely, Physio-Synergistic Intelligence and Socio-Synergistic Intelligence.

Our developmental cognitive model would account for (but not limited to) the following cognitive functionalities:

1. Acquisition of brain internal representation of self body through whole body sensory-motor learning,
2. Learning about objects and tools,
3. Emergence and development of behavior imitation,
4. Concept/imagery formation by abstraction of behaviors, and
5. Fusing together the behavior concepts/imitation with interactions through gestures, vocalizations, and facial expressions.

Some of our models will be based on the neuroscientific models provided by the Synergistic Intelligence Mechanisms group; The model of object manipulation for topic 2, imitation learning for topic 3, and motor imagery for topic 4.

Two of the hardware platforms required for the above research are; (a) whole-body sensory-motor learning robot, and (b) behavior imitation learning robot. We will develop these robots and use them for the above described research on cognitive models. Robot (a) will have a whole body tactile sensing skin and whole body motion capabilities, and will be used for experiments such as body image acquisition by sensory integration learning. Robot (b) will emphasize object manipulation capabilities and their integration with audio-visual sensing, and will be used for experiments on imitation learning, concept acquisition, and fusing with communication. In addition, we will construct (c) a computer simulation of sensory-motor learning of fetuses, in order to carry out learning experiments and model validation for very early stages of sensory-motor learning.

2.3 Socially Syngenetic intelligence

This group studies developmental mechanisms of communication and evaluations of intelligence based on human subjects and the human society. We, humans, anthropomorphize the object for conversation. This is an essential human ability and a hint for investigating principles of communicative and intelligent machines by engineering and scientific approaches. This idea guides us to a new cross-interdisciplinary framework called Android Science [4]. In this framework, we tackle the following issues

- Development of Man-Made Man (M3: M cube): In order to realize robots that can be accepted by the human society, we need to develop a human-like robot. The robot developed in this project is different from other robot-like humanoids and very humanlike androids (??). The robot M3 is going to have about 50 pneumatic actuators, soft skin covering the whole body, and various humanlike sensors with a child size.
- Mechanisms for establishing social relationships with humans: We, humans, have various interactive behaviors for establishing fundamental relationships with

others as humans, for example eye movement while looking at each other, personal distance, shared attention, synchronized movement while interacting, and reactions with tactile sensation. These human behaviors are precisely measured and the numerical models are developed [5]. Then, these programmable models of the behaviors are going to be implemented into M3.

- Constructive approach for understanding social mechanisms of humans and robots: We are going to implement more complicated and social behaviors to the robot that has the fundamental relationships with humans. This is the constructive approach [3] that starts from the basic elements and gradationally implements more complex behaviors. Finally, we acquire knowledge on what is a human and what is the human society by using the develop robot.

2.4 Synergistic Intelligence Mechanism

SI-Mech in short, will conduct the following three issues:

1. study of neural correlates of interactive control mechanism between hands and objects,
2. investigation of the brain mechanism of imitation and imaging, and
3. construction of an acquisition model for fundamental functions of communication.

In study of neural correlates of interactive control mechanism between hands and objects, we will use tracing of line drawing tasks under various conditions to identify the neural circuits involved in smooth manipulation of hands and objects. We will also clarify the relationship between some motor and mental disorders and the neural substrates. Furthermore, Nagai et al. [7] indicated that patients with Williams syndrome can trace the drawings of objects, but can not copy them. To investigate this dissociation between tracing and copying, we will use functional imaging and neuropsychological methods to clarify the difference in brain regions involved in these behaviors.

On the other hand, mechanisms of prediction and imitation play important roles in acquiring functions of communication. Imazu et al. [2] studied neural correlates of tool (chopstick) use with functional neuroimaging and revealed the following finding: actual use of chopstick activates the right cerebellum, whereas pantomiming and imagery of chopstick use activate the left inferior parietal lobule. This result suggests that different neural networks are involved in actual use, and pantomime and imagery of tool use. This could explain the dissociation observed in patients with ideomotor apraxia: they unable to pantomime with tools, despite their ability to manipulate tools actually.

For the purpose of 'the investigation of the brain mechanism of imitation and imaging', we will first examine those of action in the human brain with functional MRI. Second we will make a research on both the deficit and the learning process of action imitation of autistic children who have communication deficits, modeling the relationship between action imitation and communication function. In addition, we are going to compare the aforesaid study of William's syndrome with that of the autism and clarify the essence of communication function.

Furthermore, the 'motor sequence prediction learning hypothesis' on the basis of the brain imaging studies as well as the anatomical ones has been proposed. There are two main assumptions in the hypothesis. First, it is assumed that the functions of left BA44 and BA45 in Broca's area differ from each other; left BA45 is involved in motor sequence prediction and left BA44 is implicated in motor sequence generation, interacting with the basal ganglia respectively. Second, left BA45 makes contact with the temporal lobe through long-distance association fibers to integrate semantic and syntactic information. Finally, in terms of 'the construction of an acquisition model for fundamental functions of communication', we plan to construct an acquisition model through brain imaging studies based on the 'motor sequence prediction learning hypothesis'.

3. VOWEL ACQUISITION BASED ON VISUAL AND AUDITORY MUTUAL IMITATION

A preliminary study of SI is briefly described, which attempts to model mother-infant interaction in vowel acquisition. A pioneering constructivist approach to building a robot that reproduces a developmental process of infants' vowel acquisition has been conducted by Yoshikawa et al. [13] inspired by the observation in infant study. They have constructed a mother-infant interaction model with robot learning capability and parrot-like teaching by caregiver. However, the robot has not listened his/her own voice, therefore nor actively explored more natural vowels similar to the caregiver.

The study presented in this section extends the previous work in the following manners seeking for more natural interaction. First, a lip is added to the robot to imitate the lip shape of the caregiver in order to accelerate the learning process by constraining the initial exploration area in the formant space which is a well-known sound feature space to distinguish vowels [9]. From the lip shape imitation, the initial locations of vowels are obtained. Second, the pentagon the caregiver's vowels construct in the formant space is utilized as the desired vowels for the robot by shifting this pentagon to the centroid of the initial vowel locations of the robot. Third, mutual imitation between the robot and the caregiver is introduced in order to obtain more natural vowels hypothesizing that the caregiver imitates the robot voice but unconsciously the imitated voice is close to one of his/her own vowels. Through this process, the desired positions specified at the second step are gradually shifted, expecting to more natural ones.

3.1 Vowel imitation based on mother-infant interaction model

An overview of the whole system is shown in Fig. 2 where a vocal robot with articulation and auditory functions interacts with a caregiver who shows lip shape to the robot, and both imitate their voices each other. We slightly modified the vocal robot in [13] by adding the lip structure that has two degrees of freedom corresponding to opening/closing in horizontally and vertically. The shape change of the vocal tract is controlled by four motors instead of five ones in the previous work.

The caregiver has two roles: the first one is to show the lip shape to the robot so that it can imitate the lip shape

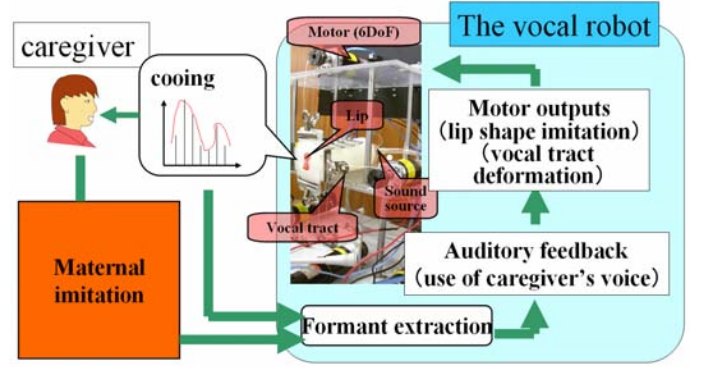


Figure 2: An overview of the whole system

since there are some evidences on the relationship between the lip shape and the vowel utterance. Carre [10] shows the formant changes owing to opening/closing of the lip by his simulation. Patterson and Werker [8] found that the infants have knowledge on the relationship between the lip shape and phoneme. Then, we assume that the robot can imitate the lip shape corresponding to each vowel from the caregiver to initialize its vowels in the formant space for further exploration.

The second one is maternal imitation, that is, imitation of the robot voices. Since the robot also imitates the caregiver's imitation, mutual imitation is expected to lead the robot voices to more natural one. The implicit assumption behind this process is that the caregiver tries to imitate the robot voices, but unconsciously his or her voices are close to his or her own vowels owing to the embodiment (sensorimotor constraint). As a result, the robot voices gradually change to more natural ones.

The vocal robot has six degrees of freedom, two of which are for lip opening/closing. First, we examine the utterance capability of the robot. The motor command is normalized into three levels 0 (free, no deformation), 0.5 (middle), and 1.0 (the maximum deformation), therefore we have 729 (3^6) utterances. Among them, we picked up 711 utterances after removing unstable ones.

Next, the effect of the lip shape is examined, that is, two of six DoFs corresponding lip opening/closing is fixed to imitate the human lip shape, and other four DoFs are combined (81 utterances, 3^4). Fig.3 shows the formant distribution of the robot utterances when the lip shape is imitated. The top row indicates the formant distributions of utterances when the lip shape is fixed to imitate that of human's vowel utterances /a/, /i/, /u/, /e/, and /o/, respectively. The bottom row shows the lip shapes of human and the robot, respectively.

The number of utterances constrained by lip shape is 394 (11 unstable utterances are removed) from 711 without the constraint. This tells that by lip shape imitation, the number of candidates for vowel utterances can be reduced to 55%.

3.2 Learning to acquire more natural vowels through mutual imitation

From the above discussion, the exploration process to find more natural vowel utterance seems easy since the robot can

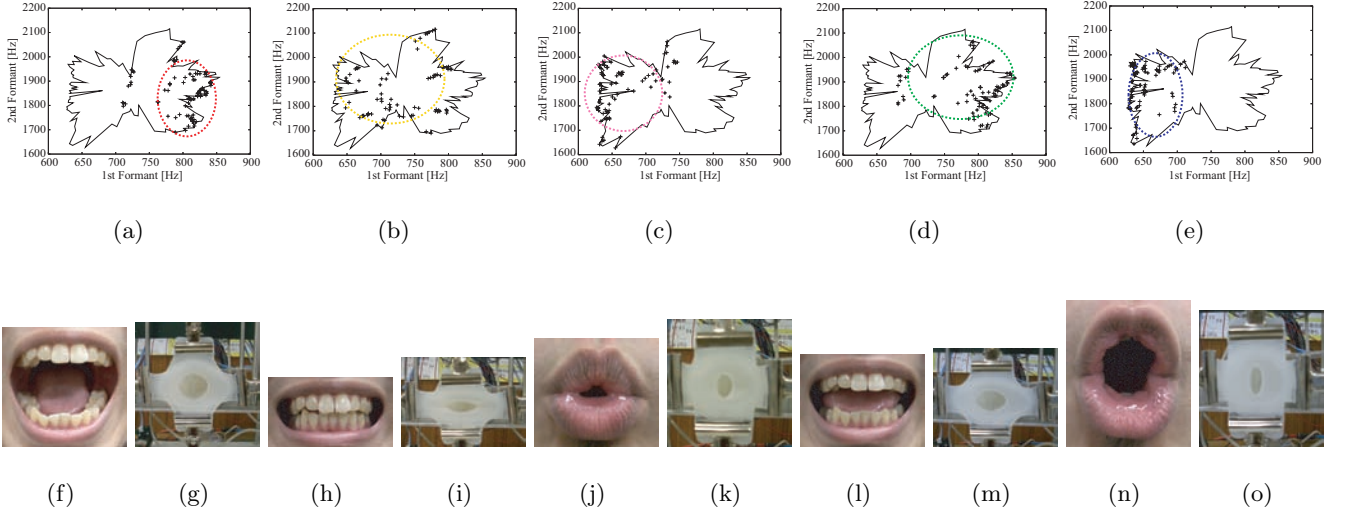


Figure 3: Formant distribution of the utterances (top: /a/, /i/, /u/, /e/, /o/) from lip shape imitation (bottom)

follow the trajectory between its initial position ($L/j/$, $j=a, i, u, e, \text{ or } o$) to the desired vowel position ($C/j/$, $j=a, i, u, e, \text{ or } o$) under the condition that the lip shape is fixed. However, there is no guarantee that the desired vowels can be heard as natural ones. Then, we introduce mutual imitation to shift the desired position so that the robot can acquire more natural ones by hypothesizing that the caregiver's imitation is biased by his/her own sensorimotor constraint, that is, embodiment, and therefore unconsciously implies the direction of natural utterance in the formant space.

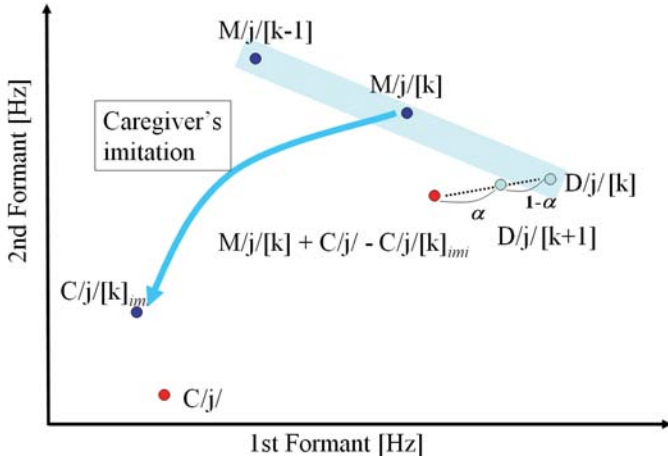


Figure 4: Mutual imitation process on the formant space

Fig.4 explains how mutual imitation can improve the robot utterance. Let $D/j/[k]$, $M/j/[k]$, $C/j/[k]_{imi}$ ($i=a, i, u, e, \text{ or } o$) be the desired position, the current position and the utterance that caregiver imitates at the time= k , respectively. Initially, $D/j/[0] = C/j/$ and $M/j/[0] = L/j/$, and $C/j/$ is

the caregiver's original vowel.

1. The robot finds the the current position $M/j/[k]$ on the trajectory between $M/j/[k-1]$ and $D/j/[k]$, and utters this position to the caregiver.
2. The caregiver imitates the robot utterance $M/j/[k]$, but the imitated utterance $C/j/[k]_{imi}$ is close to his/her own vowel $C/j/$ owing to the sensorimotor constraint.
3. A difference vector from $C/j/[k]_{imi}$ to $C/j/$ is supposed to indicate the direction to more natural vowel utterance.
4. Then, the desired position of the utterance is set at the position $M/j/[k] + C/j/ - C/j/[k]_{imi}$ from the current position.
5. To avoid a big change, a new desired position $D/j/[k+1]$ is set as follows: $D/j/[k+1] = \alpha D/j/[k] + (1-\alpha) (M/j/[k] + C/j/ - C/j/[k]_{imi})$.
6. Repeat the above process until $M/j/$ does not change (often, due to the limit of the utterance capability).

We apply the learning algorithm to the real robot and a caregiver with $\alpha = .7$. Here, we stop the learning at the 500th step since almost no changes happened in the robot utterance. Fig.5 shows the results of all vowels where thin folded lines of $D/j/$ indicate how they are modified according to the mutual imitation. The initial $D/j/s$ are shown as nodes of a pentagon connected by dotted black lines and the final ones as nodes of a pentagon connected by dotted yellow-green lines. The acquired vowels are shown as nodes of a pentagon connected by dotted sky-blue lines. The initial nodes are black and the final ones colored. The human vowels are shown as larger marks such as +, *, and so on, while the robot ones as smaller ones.

Since it is difficult to show how natural the robot utterances are, we prepared the robot utterances without mutual

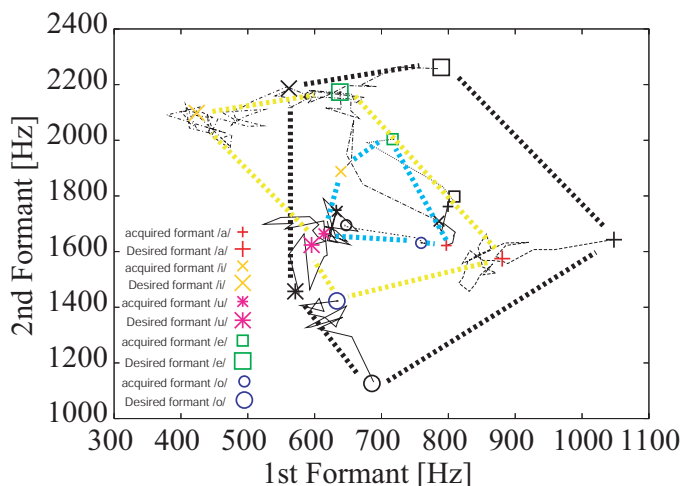


Figure 5: Desired and acquired vowels by a robot at the learning step 500

imitation, that is, the robot acquired the vowels supposing the initial desired positions D/j/[0] until the end of learning, and asked about 40 naive people (no knowledge on vowel learning). As a result, 70% agreed that the utterances with mutual imitation are more natural than that without mutual imitation.

3.3 Discussion

Currently, we emulated the lip shape imitation by hand. In a real situation, the robot should capture the caregiver's face through its TV camera and extract the lip shape information. Then, it should send the motor commands to imitate the observed lip shape. To do that, the robot should know the correspondence between the body parts such as eyes, nose, and lip, and also how to control its body part, in this case lip motion. Neonatal imitation tells us such a capability [6] though it has been controversy. However, its mechanism has not been revealed yet. Development of boy representation and motor control is one of the key issues in cognitive developmental robotics, and we will attack this problem in near future.

Another issue is how the caregiver behaviors affect the learning results. In this paper, the number of the caregiver is just one, but we have not recorded how he responded (imitated) to the robot. The analysis of such interaction data would be helpful to understand the developmental process of vowel acquisition, and also to build the design policy of the learning robots.

In our study, we adopted the following hypotheses that the lip shape information helps the infant's vowel acquisition and that the caregiver's imitation tends to similar to his/her own vowels. Based on these, we build the interaction model and showed the experimental results. The verification for these hypotheses is not complete. Collaboration with infant study and developmental psychology seems necessary not only to verify them but also to refine or modify the model.

4. CONCLUSION

The brief introduction of SI project is given and one pre-

liminary study was shown toward emergence of communication. More studies on body representation, imitation, and symbol emergence will be done in our project and hopefully, we like to build a humanoid that can show intelligent, adaptive behaviors including dynamic motions such as running and jumping, conversation skill though poor utterance, and facial expression to communicate with us.

5. ACKNOWLEDGMENTS

First of all, the author appreciates the group leaders of the project: Profs. Hiroshi Ishiguro and Koh Hosoda at Osaka University, Prof. Toshio Inui at Kyoto University, and Prof. Yasuo Kuniyoshi at University of Tokyo for the invaluable discussion with them. Next, Dr. Yuichiro Yoshikawa and Mr. Katsushi Miura are appreciated for their preparation of the paper on the preliminary study.

6. REFERENCES

- [1] M. Asada, K. F. MacDorman, H. Ishiguro, and Y. Kuniyoshi. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous System*, 37:185–193, 2001.
- [2] S. Imazu, T. Sugio, S. Tanaka, and T. Inui. Differences between actual and imagined usage of chopsticks. *An fMRI study. Cortex*, 92(11):in press, 2005.
- [3] H. Ishiguro. Toward interactive humanoid robots - a constructive approach to developing intelligent robots. In *Proc. 1st Int. Joint Conf. Autonomous Agents and Multi-Agent Systems AAMAS2002*, page Invited talk, 2002.
- [4] H. Ishiguro. Android science. In *Proc. A CogSci 2005 Workshop on Toward Social Mechanisms of Android Science*, page 25 and 26 July, 2005.
- [5] T. Kanda, H. Ishiguro, M. Imai, and T. Ono. Development and evaluation of interactive humanoid robots. *Proceedings of the IEEE*, 92(11):1839–1850, 2004.
- [6] A. N. Meltzoff and M. K. Moore. Explaining facial imitation: A theoretical model. *Early Development and Parenting*, pages 179–192, 1997.
- [7] C. Nagai, M. Iwata, R. Matsuoka, and M. Kato. Visual recognition disabilities in williams syndrome - why can they trace but cannot copy? -. *Japanese Journal of Neuropsychology*, 17:36–44, 2001.
- [8] M. L. Patterson and J. F. Werker. Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6(2):191–196, 2003.
- [9] R. K. Potter and J. C. Steinberg. Toward the specification of speech. *Journal of the Acoustical Society of America*, 22:807–820, 1950.
- [10] C. R. Prediction of vowel systems using a deductive approach. In *In Proceedings of the International Conference on Spoken Language Processing 96*, pages 434–437, 1996.
- [11] K. H. T. Takuma and M. Ishikawa. Design and control of a 3d biped robot actuated by antagonistic pairs of pneumatic muscles, Sep. 2005.
- [12] K. Terada, Y. Ohmura, and Y. Kuniyoshi. Analysis and control of whole body dynamic humanoid motion - towards experiments on a roll-and-rise motion. In *Proc. of IEEE/RSJ International Conference on*

Intelligent Robots and Systems 2003 (IROS '03),
pages (CD-ROM), 2003.

- [13] Y. Yoshikawa, M. Asada, K. Hosoda, and J. Koga. A
constructivist approach to infants' vowel acquisition
through mother-infant interaction. *Connection
Science*, 15(4):245–258, Dec 2003.