

# A Study of Constructing a Thinking Process Model Based on Multimodal Behavior Analysis

**Shogo Ishikawa**

Shizuoka University

3-5-1 Johoku Nakaku Hamamatsu Shizuoka Japan

shogo@pooh.cs.inf.shizuoka.ac.jp

81-53-478-1472

**Shigeyoshi Kitazawa**

Shizuoka University

3-5-1 Johoku Nakaku Hamamatsu Shizuoka Japan

shogo@pooh.cs.inf.shizuoka.ac.jp

81-53-478-1471

**Shinya Kiriyama**

Shizuoka University

3-5-1 Johoku Nakaku Hamamatsu Shizuoka Japan

kiriyama@inf.shizuoka.ac.jp

81-53-478-1472

**Yoichi Takebayashi**

Shizuoka University

3-5-1 Johoku Nakaku Hamamatsu Shizuoka Japan

kiriyama@inf.shizuoka.ac.jp

81-53-478-1486

## ABSTRACT

Our goal is to represent commonsense knowledge as computational models, which are applied to spoken dialogue systems that realize smart man-machine communications by correctly understanding speakers' intentions and emotions. For this purpose, we have constructed a multimodal speech behavior corpus which includes metadata annotated from various viewpoints in everyday life. This paper describes a methodology of modeling of thinking processes in problem solving based on child development, by analyzing data of attention-catching.

## Author Keywords

Multimodal behavior analysis, corpus, thinking process model, attention-catching.

## ACM Classification Keywords

H.1.2 User/Machine Systems : Human factors.

## INTRODUCTION

In the past researches on information technology, there has been much progress in expert systems which work for the purpose of limited domains. Significant though such progress is, there has arisen strong need in recent years for the development of systems which support human activities in daily life. However it is required a framework, which systems can communicate human in same thinking level, to

realize useful human-computer interaction [1].

Our intention is establishing a methodology for elucidation of thinking and developing Human Interfaces which understand speaker's intention and properly respond to situation demands by modeling complex thinking processes. For this purpose, we have constructed a multimodal speech behavior corpus which includes metadata annotated from various viewpoints, such as: utterances, actions, emotions and thinking, for analyzing behavioral factors in thinking processes from various perspectives in everyday life. As related research, Deb Roy's group is collecting child behavior video data from 0 to 3 years old children [2]. They aim to develop a computational framework where referential and functional meanings are modeled simultaneously. Their approach, however, depends on existence of natural language processing models. We have performed a multimodal analysis based on an original corpus. Since children are naive, their thinking tend to be expressed in their behavior. We aim not to obtain the findings in children development, but to construct thinking models for developing advanced situation understanding framework, using the examples about children development. In this paper, we will describe how the methodology of analysis based on the corpus enabled us to make a detailed analysis of child development and advance elucidation of thinking process by comprehensive modeling.

## AN ENVIRONMENT FOR MULTIMODAL ANALYSIS OF CHILD BEHAVIOR

### Child Learning Environment

We have experimental parent-child learning environment [3]. Three sixty-minute classes are held weekly, each class consisting of three parent-child pairs, where children are of the same age. From June 2005 to December 2009, we have held 327 classes for months, and recorded 496 hours of video.

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, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IUI 2010, February 7–10, 2010, Hong Kong, China.

Copyright 2009 ACM 978-1-60558-246-7/08/04...\$5.00

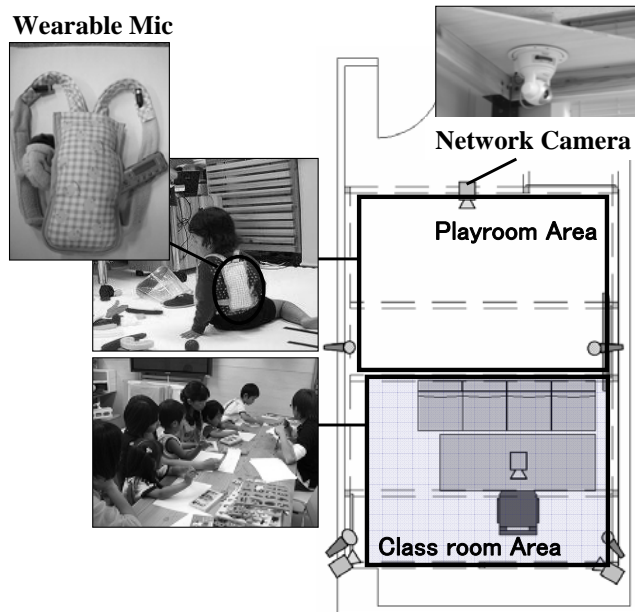


Figure 1. Children learning environment.

In order to observe child's behavior, we place four cameras which allow us to operate them by remote control. The new device allowed remote operations of pan, tilt, or zoom via the internet. These operations gave us the way to cope dynamically with blind spots caused by active movements of children. We developed a wearable speech recording device. Two condenser microphones were arranged near both shoulders. Recorded speech was stocked in a voice recorder stored inside of a rucksack. The new device drastically enriched the quality of recorded speech; the noise reduction of 17 dB was achieved, and speaker segmentation using the recorded data became possible [4]. Utilizing the speech data recorded by the developed device enables us to analyze child's utterances in great detail. Figure 1 illustrates the environment of child learning school.

### A Multimodal Speech Behavior Corpus

In this study, we analyze children's behavior based on a multimodal speech behavior corpus. The corpus consists of video, audio and text data accumulated through analyzing interactions between children, parents, teachers and objects in the child learning school. This data was annotated with metadata on external features (e.g., utterances, gestures, direction of eye gaze) and internal features (e.g., emotions, thinking, intention). The corpus enables us to analyze how social development results from natural and spontaneous behavior in everyday life. We put a strong focus on the process of objectivization by investigating the hypotheses generated through subjective observation and insight from multiple viewpoints. The interpretation of mental description concerning human mind is diverse. The corpus has a flexible schema that experts, supervisors and users from related fields generate their hypotheses by considering

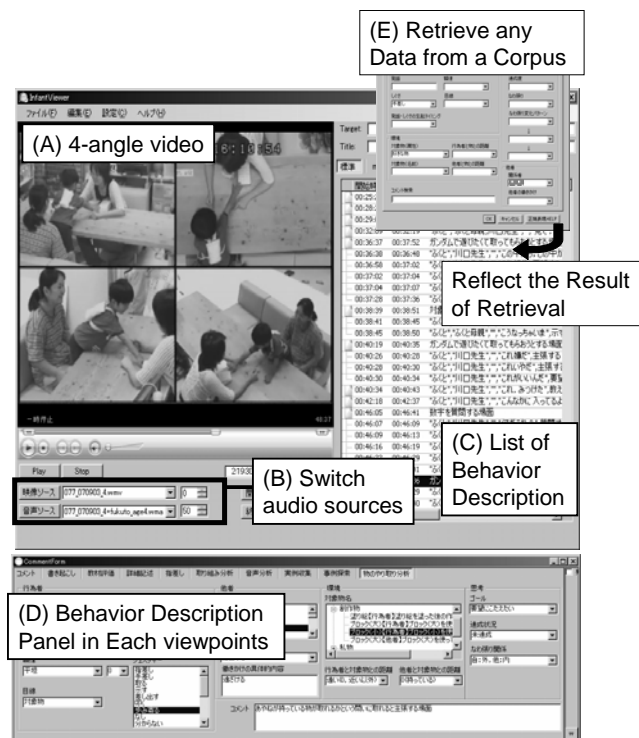


Figure 2. A multimodal behavior analyses tool.

similar scenes related to the view point extracted from the corpus. As a result, it enables us in-depth relative analysis of developmental factors.

### A Design of a Tool for Behavioral Analysis

In order to facilitate analysis of behavior in multiple viewpoints, we have developed a support tool for describing multimodal interaction, and retrieving various descriptions (see Figure 2). This tool has functionality from switching between target child audio sources (Figure 2 (B)) while watching multi-angle video (Figure 2 (A)). This enables the analysis of each child to be carried out easily. It also has a function for displaying a list of annotated data and behavior description data, which annotated in the past, for comparing scenes of the similar situation (Figure 2 (C)). For tagging the metadata from various viewpoints, it supplies selectable description mode for accumulating structured descriptions according to specific viewpoints (Figure 2 (D)). A retrieval function allows analyzing how various descriptions relate to each other, and extracting scenes similar to what we want to analyze from the corpus (Figure 2 (E)).

## DEVELOPMENTAL ANALYSIS FOCUSED ON ATTENTION-CATCHING

### Multimodal child Behavior Annotations

We construct a behavior model focused on situation when a person catches other person's attention. Behavior for catching attention is an essential component in communication. In human-computer interaction, the ability

to catch one another's attention seems to be an indispensable factor.

For analyzing catching-attention, we used physical expressions data in the behavioral corpus for analyzing ways children develop to catch the attention of others [5]. This data contains annotations on six scales: age, contents of utterances, prosodies, gazing directions, gestures and intentions. The data is on 240 scenes collected from observing a child through a 10-month period (14 to 23 months old).

Figure 3 shows an example of a scene in which K incurs behavior towards a goal, "I want to give the object I'm holding to the teacher". This is an example of how we relate a description of physical properties to a description of mental properties in the corpus. We annotate results of analyses about thinking, which is a factor of behavior, as metadata based on description data of behavior from the perspective of the child. Through comparing a connection between descriptions of physical expression skills and mental descriptions each age in month using this annotated data, we formulate hypotheses of mental development.

### Hypothesis formulation of mental development

We analyze mental factors behind attention-catching by discussing described data as showed in previous section between multiple participants. Our analysis resulted in formulating a hypothesis of mental development was formulated.

The results showed that it is represented catch attention as following four-stage process of mental development as

K's Goal : I want to give the object I'm holding to the teacher				
Video	(A-1)	(A-2)	(A-3)	(A-4)
Description items				
Situation	K tries to give a teacher his holding object	K tries one more time.	K tries yet another time.	K gives a teacher his holding object
Utterance	Dozo (Here it is)	Dozo (Here it is)	Dozo (Here it is)	Dozo (Here it is)
Prosody	Normal (flat F0 and intensity)	Emphasis (high average and rising intensity)	Emphasis	Normal
Direction of eye gaze	An object	An object → A teacher	A teacher	A teacher
Gesture	Pass	Pass	Pass	Pass
Annotate mental descriptions ↑				
Attention	K pays attention to an object	K changes attentional target	K pays attention to a teacher	K pays attention to a teacher
Behavioral factor	K Shows his desire	He changes a teacher's situation	He changes a teacher's situation	If a teacher changes her situation, K do.

Figure 3. An example description of a scene where K gives the object he is holding to a teacher. He ends up trying several times because the teacher is not aware of him (A-1, A-2, A-3), and finally succeeds when the teacher has become aware of his intention (A-4).

shown in Figure 4.

- **Instinct:** The child has insufficient representing ability, his focus doesn't settle on objects, and unit time of behavior is short. He mainly reacts instinctively when he sees an object.
- **Concentration:** The child has a wider range of expression, is able to look at a single object for a long time, and is able to maintain his mental condition to achieve his goal. However, he isn't capable of fully considering the other persons' situation.
- **Trial and Error:** The child starts to pay attention to others' needs. He learns to make the link between different physical expression skills and other persons' situation through dealing with various situations.
- **Situation Understanding:** the child considers the other persons' situation first and he behaves by selecting from learned skills depending on their situation contexts.

The developmental process (Figure 4) is linked to represent process of learning prosocial behavior concurrently with improvement of attention ability. Hence, the finding of the connection between physical expression skills and mental factors shows that our method of multimodal behavior analysis is effective.

### Interpretation of Developmental Process of Attention-catching

The development of strategies for attention-catching represent a change in what the child is conscious of when about to achieve a goal. That development can be

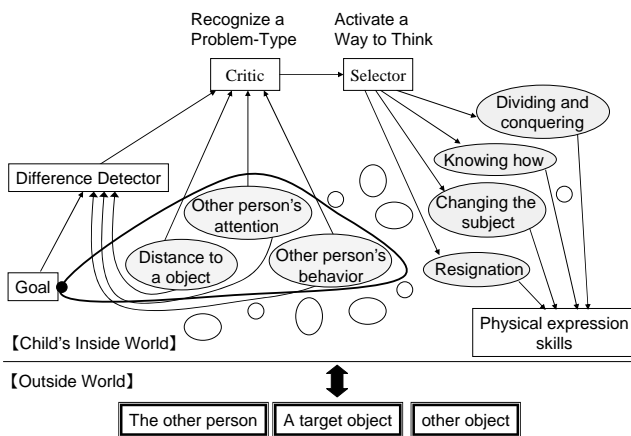
Level	External feature	Internal feature
Situation Understanding 21 month ~	<b>[DEG]</b> Look at other person <b>[DEG, U]</b> Wait for just the right time <b>[U, G]</b> Not natural behavior	<b>Understand</b> situation to notice <b>[Consider other person's situation]</b>
Trial and Error 18 month ~	<b>[DEG]</b> Look at both an object and other person <b>[U, G]</b> Change depending on situation <b>[DEG, G]</b> Check other person carefully	<b>attention allocation</b> <b>[Observe other person]</b>
Concentration 16 month ~	<b>[DEG]</b> Look at an object <b>[DEG, G]</b> Try many times <b>[DEG, U, G]</b> Unit of an action is long	<b>Continually</b> paying attention to an object <b>[obsession with an object]</b>
Instinct 14 month ~	<b>[DEG]</b> With poor aim <b>[DEG, U, G]</b> Unit of an action is short	<b>scattering</b> of attention <b>[instantaneous transition of interest]</b>

Figure 4. Developmental changes of catch attention. U: utterance, DEG: direction of eye gaze, G: gesture..

interpreted in two ways.

One possibility is that the development takes place as a result of increased ability to pay attention to ones surroundings. Especially, transition from instinct to concentration stage significantly improves the attention ability of a child. This is associated with mechanism of thinking where the attention ability involves ability of controlling short-term memory [6]. As mentioned above, the development process is able to take change of ability, which it maintains state of permanent mental activation of target subject.

The other interpretation is that change from selfish approach to social approach as a result of natural growth. In the interest of joint attention that deeply relates catch attention, Bruner indicates that it needs to joint attention in



**Figure 5. An attention-catching model. First, objects to be recognized in the world are determined by the child's goal. Next, The Critic recognizes a problem type by analyzing differences between the goal and current state of recognition items. Then, after recognizing the problem type, the Critic activates a Selector. The Selector then scans for a proper solution and produces behavior visible to the outside world.**

	Situation Recognition	Critic	Selector
<b>Instinct</b>	·A target object ·Other object	Unsated desire	Use a usual way
<b>Concentration</b>	·A target object ·Other person's behavior	An action didn't achieve	Find a better way to predict
<b>Trial and Error</b>	·A target object ·Other person's behavior ·Other person's situation	I've repeated the same thing	Some other process must be stuck
<b>Situation Understanding</b>	·Other person's behavior ·Other person's situation	Conscious of other person	Do social behavior

**Table 1. Thinking factors in each developmental level.**

social interaction [7]. He found that the developmental change is a process of acquiring the skills which effectively represent joint attention in communication with others. In the interest of ways of thinking, it was found that the developmental process is able to interpret the multi-level thinking model proposed by Minsky [8]. His model represents change from instinctive reaction to social behavior through experience.

Examining the developmental process of attention-catching can be used as a foundation for hypothesis based on the human thinking model, and has excited new insights in related research fields such as, developmental psychology and brain science.

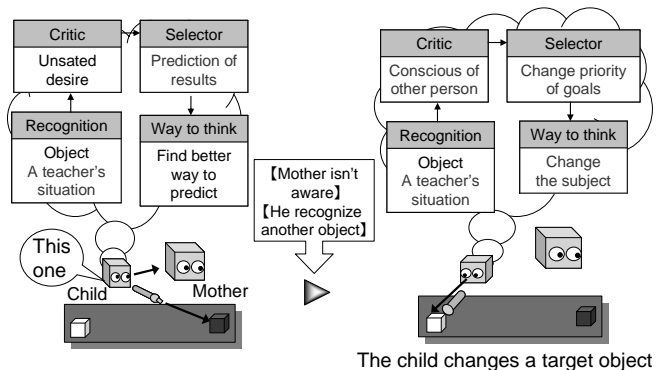
**REPRESENTATION OF THINKING PROCESS BY AN ATTENTION-CATCHING MODEL**

In this section, we propose a model which represents thinking processes involved in attention-catching behavior. It was showed that the model is effective for new kinds of analyzing of thinking.

**Design of an Attention-catching Model**

Based on the hypothesis of developmental process that was mentioned in the previous chapter, we designed a behavioral model of attention-catching, which is used to represent phases of a thinking process. We focused on scenes where a child tries to get a distant object by catching another person's attention. We made the situation design simple by limiting the model to only relating to four objects, child, the other person, a target object and other object which the child doesn't want to. Under the preconditions, we employed the Critic-Selector model proposed by Minsky [8] in designing the representation model. The Critic-Selector model represents a basic structure of human problem solving. When a person is faced with a problem, the Critic part of a reasoning process infers the problem type. In the Critic-Selector model, the reasoning process for problem solving is divided into two parts. The Selector part activates proper ways of think utilizing the person's knowledge and memories. Figure 5 is an illustration of the attention-catching model. Table 1 shows thinking processes in each developmental level.

Figure 6 represents the situation that a child (22 month-old)



**Figure 6. Representing a actual scene by the model.**

wants to get a distant target object, but he gets another object because his mother is unaware. Since he gives up achieving his goal by considering his mother's situation, he thinks in social thinking process where he understands other person's situation. The model bring us understanding thinking process.

Our model is unique for representing complex thinking process. We represent attention ability, conscious of others, problem solving and physical expression skill as structurally related to each other. Although focused on a limited range of scenes, the model represents the problem solving process in an abstract form. Through abstraction, the model is able to represent a scene with various recognition items, problem types and solutions, if the goal is that the child gives his object others in attention-catching.

### Discussion

Analyzing behavior by the attention-catching model captures thinking work, we were able to link processes concerned with physical expression skills, attention ability, understanding others, and problem solving ability in each developmental level. However, it seems there is hardly representation of scenes by the model. For example, it includes systems of memories, which return to a past subject after forced interrupt, and influence of other person's model (distinction between a mother and another people). These scenes are helpful factors for the sophistication of the model and demonstrate the relationship to other aspects. Our methodology for analyzing behavior has flexible structure permitting a new viewpoint. Therefore, it is able to analyze deep-level thinking process by reformulating description items.

It indicates that the model has the potential for applying an interactive system in the following two respects. The first is how to use a way of solving a problem. The system runs proper problem-solving depending on the situations. The other is smooth communication. In collaborative work, not only human but also the system need to know the way of thinking with each other. When the system understands human's situation and outputs the system's way of thinking, we are able to communicate with it. Consequently, the model contributes to useful human-computer interaction in the future.

### CONCLUSION

We proposed a methodology for explicit analysis of development process and deep-level modeling of thinking. This methodology makes connections between behavior and thinking in real-life situations by multimodal behavior analysis based on a corpus. We found that the methodology

is effective for research concerned with complex thinking analysis, and it enabled us to carry out behavior analysis, hypotheses formulation and representation of thinking models consistently. The attention-catching model based on developmental analyses, we were able to associatively link processes concerned with physical expression skills, attention ability, understanding others, and problem solving ability in a single system. Our model does therefore make a contribution to the understanding of mental aspects related to thinking process.

In the future, we plan to further analyzing of thinking process by refining the model based on feedback obtained from simulation approach in domain specific situation, such as collaborative work.

### ACKNOWLEDGMENTS

This research was supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) of Ministry of Internal Affairs and Communications in Japan.

### REFERENCES

1. Norman, D.A. *The Design of Future Things*. Basic Books, 2007.
2. Roy D., et al. The Human Speechome Project. *Proc. of the 28th Annual Cognitive Science Conference*, 2006.
3. Kiriyama S., et al. A Large-scale Behavior Corpus Including Multi-Angle Video Data for Observing Infants' Long-term Developmental Processes. *the Ninth International Conference on Multimodal Interfaces (ICMI2007)* (2007) 186-192.
4. Tsuji R., et al. Observations of the Spoken Language Acquisition Process Based on a Multimodal Infant Behavior Corpus. *Interspeech2006* (2006) 1356-1359.
5. Ishikawa S., et al. A Study of Infant Development in Demonstrative Expression Based on Multimodal Behavior Observation. *International Symposium on Skill Science 2007* (2007) 87-92.
6. Minsky, M. *The Society of Mind*. Simon & Schuster, 1988.
7. Bruner, J. From Joint Attention to the Meeting of Minds: An Introduction. In Moore, C & Dunham, P.J. (eds.). *Joint attention; Its origin and role in development*, Lawrence Erlbaum Associates (1995) 1-14.
8. Minsky M. *The Emotion Machine*. Simon & Schuster, 2006.