How Children Acquire Situation Understanding Skills?:
A Developmental Analysis Utilizing Multimodal Speech Behavior Corpus

Shogo Ishikawa¹, Shinya Kiriyama², Yoichi Takebayashi¹, Shigeyoshi Kitazawa²
¹Graduate School of Science and Technology, Shizuoka University, Japan
²Faculty of Informatics, Shizuoka University, Japan
(shogo@pooh.cs., kjriyama@, takebay@, kitazawa@)inf.shizuoka.ac.jp

Abstract
We have developed a multimodal speech behavior corpus which includes metadata annotated from various viewpoints such as, utterances, actions, emotions and intention for analyzing behavioral factors of thinking processes from various perspectives in everyday life. Utilizing the corpus, we analyzed child development of situation understanding skills focused on “attention-catch” that has a role as a signal when communicating with other people. We formulated a hypothesis of the developmental process that there is a connection between physical expression skills and mental conditions such as utterances, gestures and the attention ability. The analysis results showed that the situation understanding skills follow the similar development, which is a change of object-centric to person-centric, despite the age of developmental change is different. Furthermore, the analysis results provided us with a more in-depth construction of the corpus.

Index Terms: multimodal speech behavior corpus, infant development, behavior analysis, situation understanding skills.

1. Introduction
Situation understanding skills are an essential component of spoken language acquisition. Based on situation understanding skills, people understand multiple modalities, such as utterances, gestures and emotion, and use proper modalities depending on the situation. Each modality has its own specific functions, and their fundamental properties are acquired in childhood. Tomasello suggests that situation understanding skills are related to socialization of attention and understanding of other person’s intention [1].

The progress of corpus has contributed to the research field of language acquisition and cognitive development. The LENA Foundation has been developing large-scale corpus and advanced technology for treatment of language delays and disorders [2]. The ACORNS project aims to develop a computational framework to create an artificial agent that is capable of acquiring human verbal communication [3]. Corpus for observation of child behavior have been evolving into large-scale and high-capacity ones, since the progress of their development technologies has made them multilingual and multimodal. However, the child development is a complex mix of mechanism and behavioral factors. Therefore, it needs to develop a corpus for deep-level analysis of the child development.

We have constructed a multimodal speech behavior corpus that has annotated metadata such as, utterance, gesture, prosody emotion and ownership for child behavior in everyday life. It forms a basis for analyzing mental condition. We have performed the multimodal analysis utilizing our corpus and design of our corpus [4, 5]. In this paper, we describe how our corpus enabled us to make in-depth analysis of the child development.

2. A method for analyzing behavior based on the multimodal speech behavior corpus
We utilized a multimodal speech behavior corpus for analyzing child development by accumulating metadata of external and internal features. The corpus consists of video, audio and text data accumulated through analyzing interactions between children, parents, teachers and objects in a child learning school. From June 2005 to February 2010, we have held 331 classes, and recorded 505 hours of video. The corpus enables us to analyze how social development results from natural and spontaneous behavior in everyday life. Starting with utterances and other external modalities such as gestures and the direction of eye gaze, we generate possible hypotheses of child development. Especially, since utterances are obvious data and visualized, they are an important clue for mental analysis. For a better quality of speech data, we have developed a wearable speech recording device, where the noise level was improved by 17 dB in comparison to a microphone embedded in the environment [4].

Fig. 1 shows evolution of the relationship between speech corpus and behavior analysis research. Corpus for analysis of child behavior has been expanding into large-scale and high-capacity ones, since the progress of their development technologies has made them multimodal. Simultaneously, the behavior analysis has accumulated findings of internal features, such as emotion and intention, from multiple viewpoints. We need to
Table 1: Metadata for describing child behavior.

<table>
<thead>
<tr>
<th>Viewpoints</th>
<th>Description items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical</td>
<td>utterances, prosodies, gestures, direction of eye gaze</td>
</tr>
<tr>
<td>expressions</td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>other person, a target object, distance between a actor and a object,</td>
</tr>
<tr>
<td></td>
<td>distance between other person and a object, other person’s reaction.</td>
</tr>
<tr>
<td>Goal</td>
<td>child’s goal, e.g. catch teacher’s attention, get a toy, give a toy to the mother.</td>
</tr>
<tr>
<td>Mental conditions</td>
<td>attention-catching, cognition, emotion, ownership.</td>
</tr>
</tbody>
</table>

deal with an integration of mental analysis and technologies for constructing the corpus. However, the interpretation of the human mind is diverse.

Then, we put a strong focus on the process of objectivization by investigating the hypotheses generated through subjective observation and insight from multiple viewpoints. The corpus has a flexible schema where experts, supervisors and users from related fields generate their hypotheses by considering similar scenes related to viewpoints extracted from the corpus. As a result, it enables in-depth relative analysis of developmental factors.

3. An environment of multimodal child behavior analysis

3.1. Multimodal child behavior annotation

We focused on attention-catching behaviors for analyzing situation understanding skills, because they convey basic intentions, to catch another person’s attention, to get something, and to mention something. Attention-catching behaviors are represented with rich cues, such as utterances and gestures, which are easy to observe, encourages us to focus on them.

We have designed a metadata for analyzing child developmental factors shown in Table 1. Physical expressions and Situational items are decided by the analysis of child demonstrative expressions[5]. To represent communication with others, it is necessary to capture a sequence of scenes with a consistent goal. We call it the goal-oriented approach. The unit of observation is a round of flow till accomplishing or failing a goal, when the child acts toward the goal. We annotated metadata to mental condition of child behavior by a goal-oriented approach.

3.2. Design of a tool for behavioral analysis

In order to facilitate analysis of child behavior in multiple viewpoints, we have developed a support tool for describing interactions, and retrieving various descriptions. This tool has the functionality from switching between target child’s audio sources while watching multi-angle videos. This enables the analysis of each child to be carried out easily. Fig. 2 shows an example of a scene in which a child behavior persists towards a goal: “I want to get the distant object.” It is an example of how we relate a description of physical properties to a description of mental properties in the corpus. The data consists of 240 scenes collected from describing a child through a 10-month period (14 to 23 month-old).

A retrieval function as shown in Fig. 3 allows the analysis of how various descriptions relate to each other, and the extraction of similar scenes. The function uses basic data, like a persons’ and objects’ properties (e.g. age, attribute), to generate queries. Our tool is analyzed similar scenes by narrowing down a search. By using the metadata and through comparison of the connections between descriptions of physical expression skills and mental conditions at different month periods, we are able to analyze mental development.

4. Developmental analysis of situation understanding skills utilizing the corpus

4.1. Mental analysis based on the development of physical expression skills

We analyzed child’s attention-catching behavior utilizing the corpus. Based on the developmental analysis of demonstrative expressions, physical expression skills were divided into four levels as shown in Table 2. They represent features of utterance, direction of eye gaze and gesture in each stage. Using the level of behavioral type, we analyzed child behavior in similar...
Table 2: Level of behavioral type. U: utterance, DEG: direction of eye gaze, G: gesture.

<table>
<thead>
<tr>
<th>Behavioral level</th>
<th>Physical expression skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>level 1</td>
<td>DEG: With poor aim. DEG, U, G: Unit of an action is short.</td>
</tr>
<tr>
<td>level 2</td>
<td>DEG, G: Try many times. DEG, U, G: Unit of an action is long.</td>
</tr>
<tr>
<td>level 3</td>
<td>U, G: Change depending on situation. DEG, G: Check other person carefully.</td>
</tr>
<tr>
<td>level 4</td>
<td>DEG, U: Wait for just the right time. U, G: Not natural behavior.</td>
</tr>
</tbody>
</table>

Fig. 4: Scenes where the child repeats behavior for catching mother’s attention.

Fig. 5: Frequency of behavioral level in similar scenes. Similar scenes 1 (goal: get an object, distance between the other person and the object: close, 23 scenes), similar scenes 2 (Gesture: point by finger, Teacher, distance between the child and the object: distant, 43 scenes), similar scenes 3 (Other person’s reaction: No reaction, 30 scenes).

The development of strategies for situation understanding skills represents a change in ways of understanding a situation when about to achieve a goal. The development takes place as a consequence of increased ability to pay attention to the one’s surroundings. The developmental process is associated with the mechanism of thinking where the attention ability involves the ability of controlling short-term memory [6]. As mentioned above, the development process is able to take a change of ability, which it maintains a state of permanent mental activation of a target subject. In the interest of joint attention that deeply relates to attention-catching, Bruner indicates that it needs to joint attention in social interaction [7]. The developmental change is a process of acquiring the skills which effectively represent joint attention in communication with others.

The developmental process is linked to represent the process of learning prosocial behavior concurrently with the improvement of the situation understanding skills. Hence, the finding of the connection between physical expression skills and
mental factors showed that our corpus is effective for in-depth analysis.

4.3. Evaluation

In order to evaluate developmental processes, we conducted developmental analysis of other children. Two boys and a girl of the same child’s age were compared to the target child. The scenes were extracted from the corpus by items and descriptions. They include “Other person’s reaction: No reaction” (similar scenes 3) as we mentioned in Section 4.1. Table 3 shows a breakdown of extracted scenes in each age period. The results are shown in Fig. 7. There are few instinct levels, and the age of change is different. H.K behaves at a “trial and error level” when 14 month-old, and early behaves at “considering other person level”. However, the result indicates that each child changes from “object-centric” to “person-centric”. Namely, children’s situation understanding skills follow the similar development process despite the age of developmental change is different.

Otherwise, there are some scenes where are hardly explained by description items. For example, S.T catches her mother’s attention by taking directly her mother’s hand and leading her to a target object. It becomes a new metadata to be annotated and a clue to analyze common features of mental conditions. We continue to analyze other children’s behavior for designing metadata.

5. Conclusions

We analyzed child development of situation understanding skills based on the multimodal speech behavior corpus. We found that the analysis utilizing our corpus enabled us to carry out explicit behavior analysis and representation of child’s mental development consistently. The developmental analysis of other children showed that situation understanding skills change similarly from object-centric to person-centric for each child. As the results of analysis, our corpus has the ability to represent in relation to physical expression skills and mental conditions, such as attention ability and understanding others. Our corpus led to be deep-level analysis of the child development. Analyzing other children’s behavior provided us with a more in-depth construction of the corpus. In the future, we plan to grow the corpus by further analysis of the situation understanding skills and by linking different viewpoints.

6. Acknowledgements

We would like to thank Hiroaki Horiuchi and Alexander Cossio for giving valuable comments on the manuscript.

7. References