2024 Joint 13th International Conference on Soft Computing and Intelligent Systems and 25th International Symposium on Advanced Intelligent Systems (SCIS&ISIS) | 979-8-3503-7333-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/SCISISIS61014.2024.10759905

Evaluation of a Writing Skill Enhancement System Through Dialogue for Children with Developmental Disorders Using Companion Robots

Sumire Hayashi
Graduate School of Engineering
University of Fukui
Fukui, Japan
hayashi.sumire@ir.his.u-fukui.ac.jp

Yasutake Takahashi
Faculty of Engineering
University of Fukui
Fukui, Japan
yasutake@ir.his.u-fukui.ac.jp

Satoki Tsuichihara
Faculty of Engineering
University of Fukui
Fukui, Japan
satoki-t@u-fukui.ac.jp

***Abstract*—We have developed a writing skill enhancement system through dialogue for children with developmental disorders and investigated the impact of the presence or absence of a companion robot and differences in input methods. Some children with developmental disorders struggle with sentence writing skill. Although previous research has discovered the effectiveness of a robot intervention in social skill training for children with developmental disorders, few studies have conducted writing skill training for them and demonstrated its effectiveness. In this study, we aim to investigate four types of writing skill enhancement systems through dialogue with presence/absence of a companion robot and input methods (speech recognition/typing) for the effectiveness of improving the writing skills of children with developmental disorders. The dialogue system uses ChatGPT, ChatAPIs, to generate responses. This paper investigated the four developed writing skill enhancement systems targeted at university students as a preliminary experiment for testing with children with developmental disorders. Experimental results showed a tendency to elicit a large number of words from the students through the interaction with the system using speech recognition and a companion robot. In addition, there was no consistent trend between the survey results and the system that was found to be most suitable, indicating the possibility of not being able to select an effective system based only on survey results.**

***Index Terms*—writing skill enhancement, developmental disorder, robot interaction, speech recognition, typing**

I. Introduction

In recent years, with the advancement of technology, robots and ICT support devices have been utilized in various field. Among them, studies have been conducted to verify the efficacy of using robots [1] [2] and ICT support devices [3] for social skill training for children with developmental disorders.

Developmental disorders include Autism Spectrum Disorder (ASD), Learning Disorder (LD), and Attention Deficit Hyperactivity Disorder (ADHD). There are individual differences, but various disorders can lead to low academic achievement and maladaptation to daily life [4]. For example, there may be difficulties in interpersonal communication, difficulties in sustaining attention and hyperactivity, and although there is no delay in intellectual development in general, there may be difficulties in specific learning abilities (reading ability, writing ability, arithmetic ability, etc.). Therefore, training is conducted to improve and solve the difficulties encountered in daily life by children with developmental disorders, such as social skill training.

While children with developmental disorders demonstrate difficulties in interpersonal communication, they often show more interest in robots than children with typical development [5]. This is due to the consistent behavior of robots which is preferred by children with ASD, making robots frequently used in interventions in social skill training for children with developmental disorders [6]. Experiments have been conducted with various robots, and the effect of robot-assisted training has been observed even when using hand-held toy robots [7]. However, conventional communication systems had problems such as the long time required for robot responses or the inability to generate appropriate responses from robots, leading to disinterest in conversation with robots [8]. Moreover, the Wizard-of-Oz method was often used, and automation of the system was an issue [9].

Some children with developmental disorders struggle with constructing sentences. Until now, research has been conducted using robots and ICT support devices in language therapy, focusing on language disorders in children with developmental disorders. Spitale M. [10] conducted an experiment using either a robot or a virtual robot randomly in language therapy for children with developmental disorders. As a result, overall language skills improved under both conditions. Especially when using a robot rather than a virtual robot, the number of utterances and changes in gaze changed significantly. Additionally, in an experiment by Mohammed F. Safi [11], they used the virtual voice assistant application "Siri" for speech training. As a result, the number of utterances and new words increased during the session time. However, much of the past research has focused on producing words and utterances, and there has been no research on writing sentences. Some children with developmental disorders can respond to questions with words, but struggle to respond with sentences or write essays. A system is needed for such children with developmental disorders to practice creating sentences on their own.

In this study, we developed a writing skill enhancement system for the purpose of improving the sentence writing ability of children with developmental disorders. In this system, the child with a developmental disorder responds to topics and questions generated by the system, thereby continuing the conversation. Children with developmental disorders input their responses in two ways: through speech recognition or typing. Previous research has often used speech recognition input, but we are also implementing typing input in our system as the ICT support room organized by Hiratani Child Development Clinic in Fukui City is practicing typing [12]. We also expect the writing skill enhancement system to motivate the children for typing practice. We also make responses from the system with and without a companion robot, in two conditions. In total, we have developed four systems with combination of two input methods (speech recognition/typing) and presence/absence of the companion robot. We compared these to investigate which system is effective in improving the ability to write sentences. We verified how those conditions affect the written contents.

In this paper, we conducted preliminary experiments targeted at university students using four types of writing skill enhancement systems developed for children with developmental disorders.

II. Writing Skill Enhancement System

A. Overview of the System

The system uses ChatGPT, Chat APIs developed by OpenAI, to generate the response to the input from the children. The ChatGPT is supposed to generate more natural responses to the child's input. The prompt to the ChatGPT includes the states as "respond within 30 characters" and "ask a question".

There are two input methods: speech recognition and typing. With speech recognition input, we use Whisper developed by OpenAI. Whisper is a free speech recognition model. It is trained with supervised data from 680,000 hours of multilingual speech data collected from the web, allowing accurate transcription of input speech.

There are also two conditions for presence/absence of the companion robot. Figure 1 shows the developed writing skill enhancement system with the companion robot. The robot is "Stack-chan RTVer.β" shown in the upper left corner of Fig. 1. The Stack-chan is a palm-sized communication robot developed and released by Shishikawa, an engineer at RT Corporation. It mainly consists of a small microcontroller module "M5Stack."

In all four system conditions, the input contents from children with developmental disorders and the response contents from ChatGPT are displayed on the PC screen as texts. In addition, the response content from ChatGPT is output as voice, regardless of the presence or absence of Stack-chan. Text-to-Speech API from Google Cloud Platform was used for voice output.

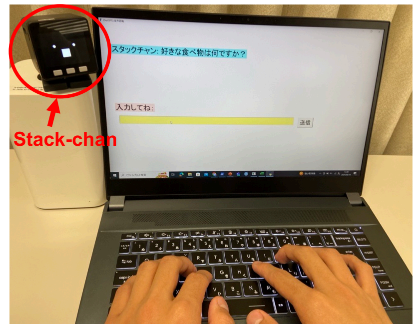

Fig. 1. System overview (input: typing, output: with robot)

B. Dialogue System Flow

The flow of the dialogue system is shown in Fig. 2. The running time of the system is 3 minutes per time. First, the system prompts the child to enter his/her name and select the first topic of the dialogue. The topics are "favorite food", "strong subject", "what I did today", "recently watched TV program", "favorite things", and "future dreams." When the child selects the topic, he/she will be taken to a dialogue screen with ChatGPT. In the dialogue, ChatGPT first asks, "What is [the selected topic] ?" The child responds to this question by speech or typing. Then the response is sent to ChatGPT and a reply to the child's input is generated. After that, the dialogue between ChatGPT and the child continues. It is also possible to change the topic in the middle of the dialogue, and if the "change the topic" button is pressed, the topic selection screen appears. The system will automatically finish the dialogue within 3 minutes after starting the dialogue.

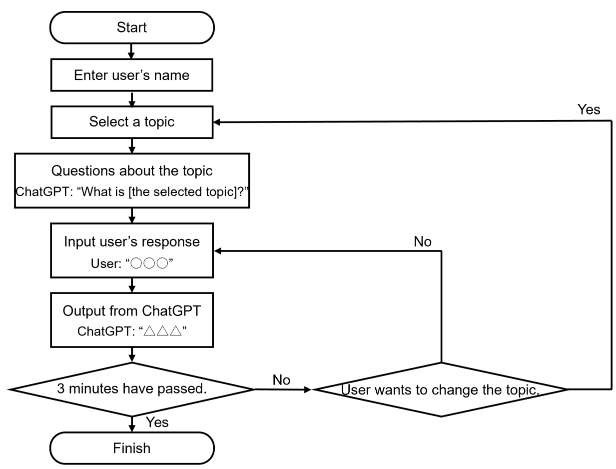

Fig. 2. Dialogue flow

III. Experiment

As a preliminary experiment to investigate which condition of the system is effective in improving the writing sentence ability of children with developmental disorders, we conducted experiments with university students, investigating the impact of input methods and presence of the companion robot on writing sentence ability.

A. Participants

8 participants were recruited at the University of Fukui. All were students from 22 to 24 years old. Their gender was all male. They volunteered to participate in the experiment.

B. Experimental method

The participants try all four conditions in a random order. The experiment time was about 30 minutes per person. During the experiment, brain waves were measured with a simple electroencephalograph(EEG). Figure 3 shows the situation of conducting an experiment while wearing an EEG. The EEG used was a two-electrode simple EEG provided by NeuroSky, Inc. The brain waves were measured for three minutes at rest before the experiment and compared with the brain waves during the experiment.

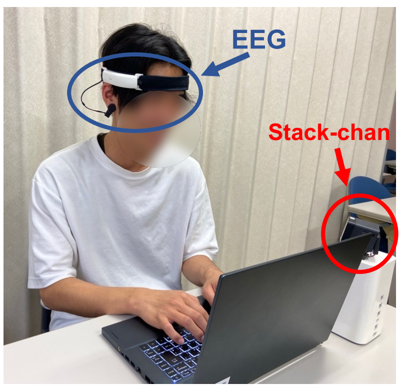

Fig. 3. Experimenting with an electroencephalograph

Also, after the experiment of each condition finished, the participants answered a survey about their impressions of the dialogue and the system. The survey consisted of ten questions: five about the impression of the conversation partner and five about the impression of using the system. This survey was answered on a seven-point scale, with 1 set as negative and 7 as positive. Additionally, after all four conditions of the system were completed, the participants answered a survey to rank the four conditions of the system.

IV. Results

A sample of the conversation during the experiment is shown in Fig. 4.

Topic : 好きな食べ物は何ですか？（What's your favorite food?）
User : 唐揚げです。（It's fried chicken.）
ChatGPT : 美味しいですよね、唐揚げ！（Fried chicken is indeed delicious!）
好きな唐揚げのお店はありますか？（Do you have a favorite place to get it?）
User : 商店街の店で食べるのが好きです。（I like eating it at a shop in the shopping street.）
ChatGPT : それは最高ですね。（That's the best!）
他に好きな食べ物はありますか？（Do you have any other favorite foods?）
...

Fig. 4. A sample interaction

From the data collected during the experiment, it was found that participants demonstrate various characteristics. The data includes dialogue data (number of utterances, number of words, time taken per response), conversation contents, and brain waves.

A. Writing Analysis

The number of written words under the interaction with the writing skill enhancement system was analyzed, and the condition that elicited the larger number of words for each participant was considered the most suitable one. As a result, "Speech/Stack-chan" was suitable for 4 people, "Speech/Control" for 1 person, "Typing/Stack-chan" for 1 person, and "Typing/Control" for 2 people. "Control" indicates the absence of the companion robot, Stack-chan. The results of the analysis of conversation data from persons suited for each condition are shown below. The contents of the participants' remarks were morphologically analyzed using MeCab [13]. Using the results of the morphological analysis, the total number of words and the number of words per second during the response time under each condition are analysed. In this case, words are counted separately as "content words", which can express meaning in one word such as nouns, verbs, and adjectives, and "function words", which cannot express meaning in one word such as particles and auxiliary verbs.

First, for those who were suited to condition "Speech/Stack-chan", the total number of words and the number of words per second during the response under each condition is shown in Figure 5. The number of words, including both content and function words, were generated the most with "Speech/Stack-chan." In addition, the average number of words generated per second during the response time was the highest when using "Speech/Stack-chan," more than seven times higher than the least "Typing/Control."

Next, for those who were suited to condition "Speech/Control," the total number of words and the number of words per second during the response time under each condition is shown in Figure 6. The highest number of words were generated with "Speech/Control," especially more function words compared to other conditions. Also, the average number of words generated per second during the response time was highest with "Speech/Control."

Then, for those who were suited to "Typing/Stack-chan," the total number of words and the number of words per second during the response time for each condition is shown in Fig. 7. The number of words was the same in "Typing/Stack-chan" and "Typing/Control", but "Typing/Stack-chan", which generated more content words, was considered suitable. On the other hand, the average number of words generated per second during the response time was higher for "Speech/Stack-chan" and "Speech/Control".

Finally, for those who were suited to "Typing/Control", the total number of words and the number of words per second during the response time under each condition is shown in Fig. 8. The number of words, including content and function words, was generated the most in "Typing/Control." However, the average number of words generated per second during the response time was the highest in "Speech/Control," and the difference among the systems was smaller compared to the people who suited other systems.

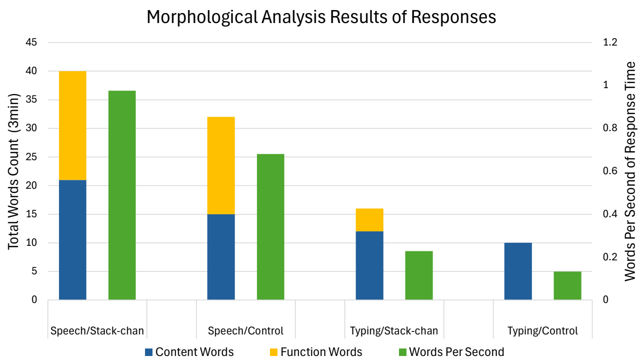

Fig. 5. Morphological Analysis Results: summary of people who best fit Condition “Speech/Stack-chan”

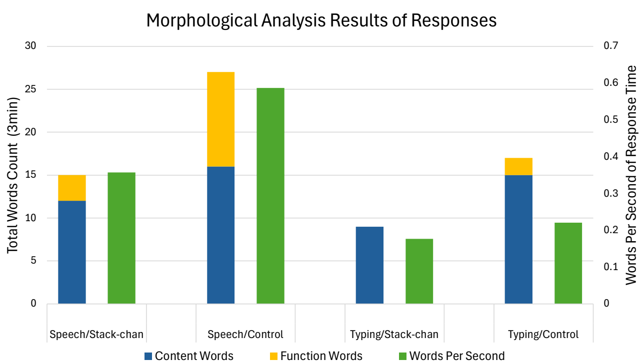

Fig. 6. Morphological Analysis Results: summary of people who best fit Condition “Speech/Control”

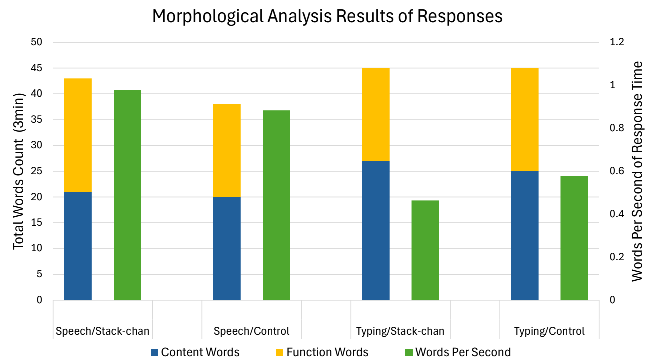

Fig. 7. Morphological Analysis Results: summary of people who best fit Condition “Typing/Stack-chan”

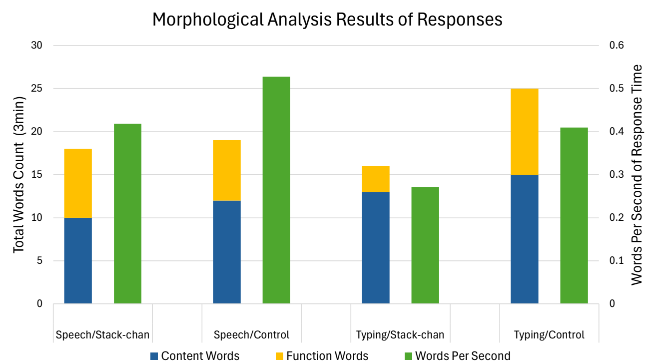

Fig. 8. Morphological Analysis Results: summary of people who best fit Condition “Typing/Control”

B. Analysis based on Brain waves

Each condition is also evaluated based on the brain waves. It is based on the measurement of whether the brain waves during the experiment exceed the average value of the brain waves at rest. As a result, 3 people had more active brains on the left, 1 person had a more active brain on the right, and 4 people had no difference in left and right. Figure 9 shows the brain waves of the person whose left brain was more active. Figure 10 shows the brain waves of the person whose right brain was more active. Figure 11 shows the brain waves of the person who had no left-right difference and found the speech input method suitable. Figure 12 shows the brain waves of the person who had no left-right difference and found the typing input method suitable. Figs. 9 and 11 show the brain waves of both the right and left were particularly active when the input was speech. In Fig. 10, the brain waves were active in the right brain when Stack-chan was accompanied. In Fig. 12, particularly in the right brain, the brainwaves were active when the input was typing.

C. Survey results

A Friedman test was performed using the results of the responses to the surveys conducted after each system. Scheffe’s paired comparisons were made between the conditions. Figure 13 shows the survey results about the impression of the conversation partner. Figure 14 shows the survey results about the impressions of using the system.

Significant differences between “Speech/Stack-chan” and “Typing/Control” were observed for “Friendliness:Unfriendly – Friendly”($p < 0.01$), “Human-likeness:Mechanical – Human-like”($p < 0.05$), “Self-awareness:Not self-aware – Self-aware”($p < 0.05$), and “Future Use:Wouldn’t Use – Would Use”($p < 0.05$). Significant differences between “Typing/Stack-chan” and “Typing/Control” were observed for “Naturalness:Unnatural – Natural”($p < 0.05$).

In addition, the results of a survey comparing and ranking the four conditions are shown in Figure 15. For the responses “I felt favorable towards my conversation partner.” “My conversation partner was approachable.” “I enjoyed the interaction.” and “I want to do it again.” many participants selected “Speech/Stack-chan” in the first place, “Typing/Stack-chan” in the second place, “Speech/Control” in the third place, and “Typing/Control” in the forth place. Similarly in the question, “It was a natural interaction”, the answers were most often in the same order, “Speech/Stack-chan” in the first place, “Typing/Stack-chan” in the second place, “Speech/Control” in the third place, and “Typing/Control” in the fourth place, although there was more variability in the responses compared

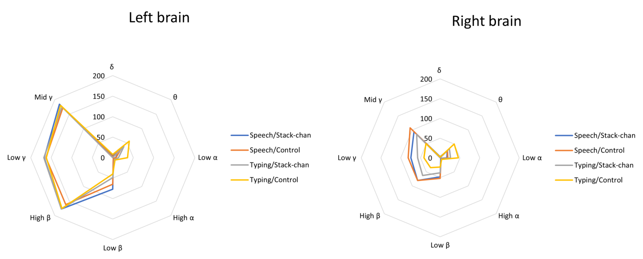

Fig. 9. Brain waves of people whose left brain is more active

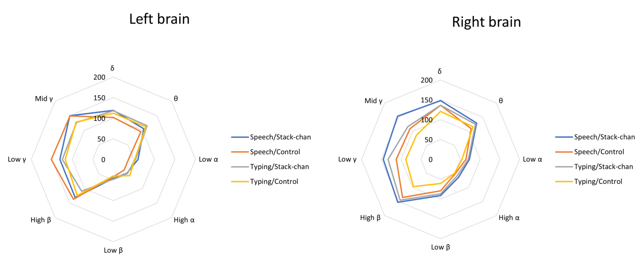

Fig. 10. Brain waves of people whose right brain is more active

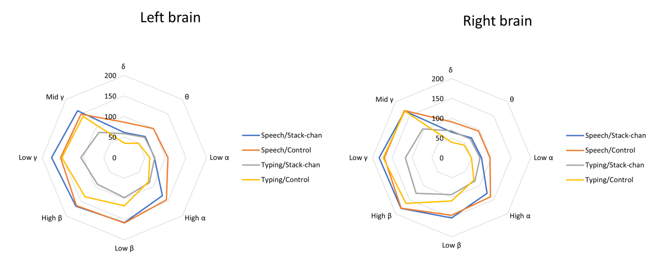

Fig. 11. No differences between left and right (Input:Speech)

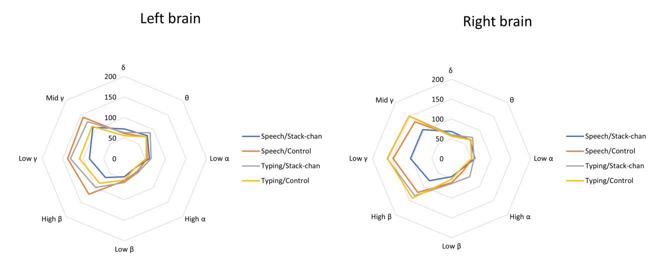

Fig. 12. No differences between left and right (Input:Typing)

to the other questions. For the question "I was nervous.", the most common answers were: "Typing/Control" in the first place, "Speech/Control" in the second place, "Typing/Stack-chan" in the third place, and "Speech/Stack-chan" in the fourth place. However, there were some who answered the opposite, with first place "Speech/Stack-chan" and fourth place "Typing/Control."

V. Discussions and Conclusions

In this study, we developed a writing skill enhancement system for children with developmental disorders. Interaction experiments were conducted to compare four conditions with for university students as a preliminary experiment to investigate the effects of the presence or absence of the companion robot and differences in input methods. As a result, it was found that the suitable input method and the effect of the presence or absence of the companion robot varied among participants.

The best system condition for the most suitable people was "Speech/Stack-chan" among four conditions. There were also people who were better suited to the other conditions. From Fig. 5, the number of written words with the best condition of the system was extremely greater than the others. From Fig. 7 and Fig. 8, it means that while the condition with the highest total word count differs from person to person, the number of words per response time is higher when the input method is speech. In addition, it was also found that as the number of words increases, the ratio of function words also increases.

It was found that there are individual differences in the characteristics of brain waves. The left brain is used for language comprehension and logical thinking, while the right brain is used when involved in image conversion and imagination. As shown in Fig. 9, people have active left brain waves for language comprehension because this system is creating sentences. From Fig. 10, because the right brain under the condition using Stack-chan was active, there was a trend for the right brain of imagination to become active due to Stack-chan. Also, in the suitable system, β waves often appear a lot, which means that you are focused when using that system.

From the "survey results", it was found that many people gave similar responses to the survey regardless of the appropriate system. From Fig. 13 and 14, "Speech/Stack-chan" was the highest rated and "Typing/Stack-chan" was often the next highest rated. Also, from Fig. 15, when comparing the four systems, "Speech/Stack-chan" was first and "Typing/Stack-chan" was second in terms of positive evaluations. From this, it was found that having Stack-chan gives a positive impression of the interaction.

From all the results, the suitable system did not match with the survey results, and those who suited any system often rated highly on "Speech / Stack-chan" and "Typing / Stack-chan" in the survey. From this, it can be inferred that having Stack-chan does not necessarily impact the number of words, but it is believed to have a positive impression on the conversation partner and a relaxing effect.

References

[1] B. Scassellati, H. Admoni, and M. Matarić, "Robots for use in autism research," *Annual Review of Biomedical Engineering*, vol. 14, pp. 275–294, 2012.

[2] N. Rakhymbayeva, A. Amirova, and A. Sandygulova, "A long-term engagement with a social robot for autism therapy," *Frontiers in Robotics and AI*, vol. 8, 6 2021.

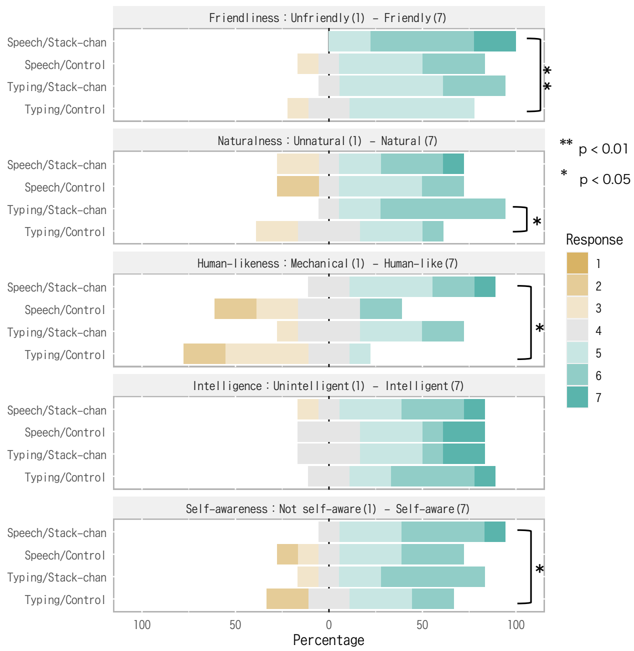

Fig. 13. Survey: Impressions of Conversation Partner

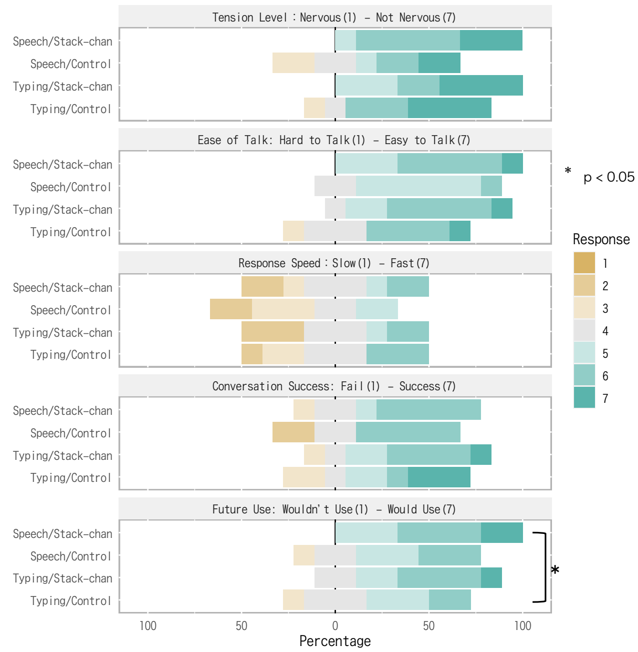

Fig. 14. Survey: Impressions of using the system

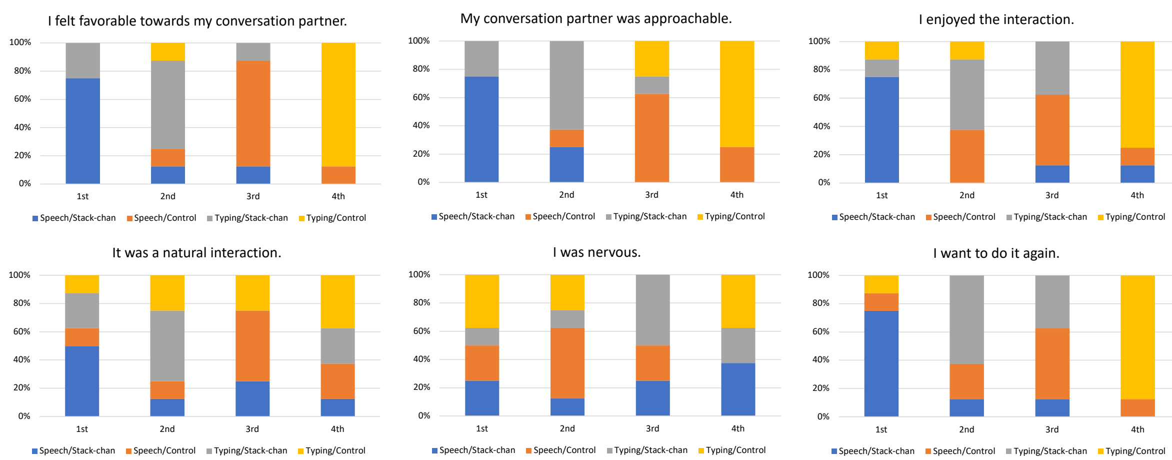

Fig. 15. The results of a survey comparing four conditions

[3] L. Bakola, N. Rizos, and A. Drigas, "Icts for emotional and social skills development for children with adhd and asd co-existence," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, p. pp. 122–131, 2019.

[4] J. Yamamoto and N. Shibuya, "Evidence-based supports for persons with developmental disabiiities: Contribution of applied behavior ana]ysis," *Japanese Journal of Behavior Analysis*, vol. 23, pp. 46–70, 2009.

[5] B. Scassellati, "How social robots will help us to diagnose, treat, and understand autism," *Robotics Research*, vol. 11, pp. 552–563, 2007.

[6] K. Dautenhahn and I. Werry, "Towards interactive robots in autism therapy," *Pragmatics & Cognition*, vol. 12, pp. 1–35, 6 2004.

[7] D. Ghiglino, P. Chevalier, F. Floris, T. Priolo, and A. Wykowska, "Follow the white robot: Efficacy of robot-assistive training for children with autism spectrum disorder," *Research in Autism Spectrum Disorders*, vol. 86, p. 101822, 2021.

[8] S. Toshiyuki, K. Takayuki, I. Michita, I. Hiroshi, H. Norihiro, and A. Yuichiro, "How quickly should communication robots respond?" *Journal of the Robotics Society of Japan*, vol. 27, pp. 87–95, 2009.

[9] T. Yoshida, "Development of a humanoid robot-based communication system for students with autism spectrum disorder," *Univercity of Fukui (master's thesis)*, 2022.

[10] M. Spitale, S. Silleresi, F. Garzotto, and M. J. Matarić, "Using socially assistive robots in speech-language therapy for children with language impairments," *International Journal of Social Robotics*, vol. 15, p. 1525–1542, 2023.

[11] M. F. Safi, B. A. Sadrani, and A. Mustafa, "Virtual voice assistant applications improved expressive verbal abilities and social interactions in children with autism spectrum disorder: a single-subject experimental study," *International Journal of Developmental Disabilities*, vol. 69, pp. 555–567, 2023.

[12] Y. Tochio, "Development of typing practice support application for learning disabilities," *Univercity of Fukui (master's thesis)*, 2023.

[13] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying conditional random fields to Japanese morphological analysis," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 2004, pp. 230–237.