



Developmental Scientific Research on Generative AI-based Dialogue System

阿部, 香織

(Degree)

博士 (学術)

(Date of Degree)

2025-09-25

(Date of Publication)

2026-09-01

(Resource Type)

doctoral thesis

(Report Number)

甲第9327号

(URL)

<https://hdl.handle.net/20.500.14094/0100498367>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



Doctoral Dissertation

Developmental Scientific Research
on Generative AI-based Dialogue System

(発達科学を取り入れた生成 AI を用いた対話システムに関する研究)

July 2025

Graduate School of System Informatics

Kobe University

阿部 香織

Abstract

In recent years, generative AI using large language models has come to make dialogue systems converse as if they were human. As the linguistic capability of dialogue systems improves, their anticipated roles are going beyond merely transmitting information, such as being long-term partners. Systems are expected to support people personally through linguistic communication. The importance of how systems interact with humans and build supportive relationships is increasing, and social adaptability and personal compatibility issues similar to those in human relationships emerge. Therefore, interactions with a generative AI-based dialogue system were investigated to improve relationships between humans and dialogue systems.

A dialogue system was developed to communicate with the subjects in natural language via voice, featuring a human-like facial expression function. These communication abilities were achieved with generative AI. The system interacted with healthy adult subjects in the communication experiment using three interaction settings. The one-to-one dyadic interaction, triadic interaction with joint attention to a topic, and non-communicative settings were designed with reference to the developmental stages of children's social communication ability. The system demonstrated its communication abilities in the dialogues, behaving differently depending on the settings. Subjective evaluations of the system and the dialogues were conducted through the questionnaire regarding the dialogue quality and the system impression. The evaluations resulted in various responses. The results were analyzed from two perspectives regarding the system: how it should interact as a social being and how it can be personally accepted by individuals.

To investigate the system's social behavior, the dialogue evaluations were discussed based on the interaction structures, that is, the positions the system took in the interactions. The communication experiment was conducted under the hypothesis that the progressive development of interaction structure in the system settings would improve the system's interaction and provide favorable dialogues for the subjects. The hypothesis was partially confirmed. The dialogues with the system received positive evaluations on the basis of well-constructed interaction structures. The system's inappropriate behavior in failed structures decreased the quality of the dialogues and worsened the evaluation of the system. The interaction structures in the system settings

needed to match the structures intended by the subjects. Under the matching and successful construction, the system fully demonstrated its dialogue capability and behaved pleasantly for the subjects. When the constructions were successful, the more advanced triadic interactions were evaluated more positively. By switching interaction structures in systems to adapt to users' demands, the structures can be successfully constructed between users and systems, and the behavior would become more appropriate for users. Socially correct behavior of systems, adapting to social contexts on partners' intention and expectation, affects users' evaluations similar to interpersonal relationships. Dialogue systems could build better relationships through the interaction in well-constructed interaction structures.

To investigate individual preferences regarding the system's behavior, the diverse subjective evaluations among the subjects from the communication experiment were discussed. The links between the evaluations of the system's impressions and the dialogue quality were analyzed for each subject using hierarchical cluster analysis. Based on the analysis results, the dialogue system's properties were classified into two types: generally and individually relational to a positive evaluation of the dialogue. The former included inspiration, a sense of security, and collaboration, while the latter included a sense of distance, personality, and seriousness. In addition, it was also shown that human-like behaviors were preferred in the dialogues. Properties of generally relational type are expected to improve dialogues for most users. Properties of individually relational type influence users' perception depending on individual preferences, and adjusting them can contribute to the personalization of systems. A design approach in accordance with individuality would be useful for making human-like dialogue systems more compatible with users.

In this thesis, ways to build better relationships between users and dialogue systems were explored. It was clarified that adjusting interaction structures is effective for adapting to social contexts, and adjusting system properties is effective for improving relationship compatibility. Appropriate and favorable behaviors of systems can be achieved with these adjustments. These findings contribute to building better relationships with advanced dialogue systems and realizing human well-being supported by future systems.

Acknowledgements

First and foremost, I would like to express my sincere appreciation to my research supervisor, Professor Zhiwei Luo, for providing me with the precious opportunity to conduct this research. I suppose I was a strange student who suddenly appeared one day. Nevertheless, I feel truly honored to have been welcomed as a member of his laboratory and to have received kind, patient, and attentive guidance. During my research in the laboratory, Associate Professor Changqin Quan and Assistant Professor Sheng Cao kindly offered suggestions and greatly supported me with their expertise. Mr. Bunka Son, who is now an assistant professor, was a cheerful companion through our doctoral student days of struggle. I would like to present my heartfelt thanks for all the support I received, which enabled me to come this far.

I would like to extend my gratitude to Professor Tetsuya Takiguchi and Professor Chikara Ohta for agreeing to review this thesis and for their valuable comments and suggestions during the degree review process.

I am grateful to my co-workers in Goyou Gakuen for their generosity in allowing me to pursue this research while working. I would like to convey my sincere appreciation to Professor Masaaki Makikawa for his guidance in the first steps of my research journey at Ritsumeikan University.

Lastly, I deeply appreciate all the participants who kindly took part in the experiments. Without their generous cooperation, any findings in this thesis would not have been possible.

Table of contents

List of figures.....	vi
----------------------	----

List of tables.....	viii
---------------------	------

1	Introduction.....	1
1.1	Dialogue Systems as Communication Partners.....	1
1.2	Human Communication Based on Language.....	2
1.3	Dialogue Systems with Human-like Linguistic Capability.....	3
1.4	Structure of This Thesis.....	5
2	Human-like Dialogue Systems in Human-Computer Interaction.....	7
2.1	Dialogue Systems.....	7
2.2	Human-like Social Interaction.....	8
2.3	Relationship Compatibility Between Users and Systems.....	12
3	Development of Generative AI-based Dialogue System.....	14
3.1	Development of the Dialogue System.....	14
3.1.1	System Overview.....	14
3.1.2	Settings of Dyadic and Triadic Interactions.....	17
3.1.3	Algorithms.....	19
3.2	Evaluation of the System and the Dialogue.....	21
3.3	Communication Experiment.....	22
3.3.1	Conditions.....	22
3.3.2	Topics.....	22
3.3.3	Subjects.....	23
3.3.4	Experimental Procedure.....	23
3.4	System Operation in Communication with the Subjects.....	26

4	Subjective Evaluation of Generative AI-based Dialogues.....	30
4.1	Methodology.....	30
4.2	Experimental Results.....	31
4.2.1	Dialogue Evaluation.....	31
4.2.2	Dialogue Evaluation from High-Rating Groups in the Dyadic and Triadic Settings.....	33
4.2.3	Dialogue System Evaluation.....	36
4.3	Discussion.....	36
5	Generative AI-based Dialogues Considering Individual Preferences.....	44
5.1	Methodology.....	44
5.2	Experimental Results.....	45
5.2.1	Questionnaire Results.....	45
5.3	Analysis and Classification.....	48
5.3.1	Analysis of Links Between Evaluations of the System and the Dialogue...	48
5.3.2	Classification of the System Evaluation Items.....	50
5.4	Discussion.....	55
6	Conclusions.....	58
6.1	Conclusions of This Thesis.....	58
6.2	Limitations and Notes.....	59
6.3	Future Work.....	60
	Bibliography.....	62
	Appendices.....	69
	Questionnaire Sheets.....	70
	Table of the Post-dialogue Questionnaire Results.....	74
	Dendrograms from Hierarchical Cluster Analysis of the Questionnaire Results.....	75

List of figures

Figure 1.1	The structure of this thesis.....	6
Figure 2.1	An example of animacy.....	10
Figure 2.2	Examples of behaviors in early childhood.....	11
Figure 3.1	An overview of the dialogue system.....	15
Figure 3.2	The face images displayed on the screen for five levels of emotional expression...	16
Figure 3.3	Triadic interaction structure between a child, a partner, and an object of interest...	17
Figure 3.4	The interaction structures between the subject, the dialogue system, and the topic, according to the settings of the dialogue system.....	18
Figure 3.5	The experimental procedure	24
Figure 3.6	An overview of the system workflow.....	25
Figure 3.7	The results for the pre-experiment questionnaire.....	27
Figure 3.8	The results for the post-dialogue questionnaire.....	28
Figure 3.9	The relationship between the dialogue evaluations and the subjects' mood after the dialogues.....	29
Figure 4.1	The results for the post-dialogue questionnaire regarding the dialogue evaluations	32
Figure 4.2	The results for the post-dialogue questionnaire regarding the dialogue evaluations in the dyadic interaction (with the subject) setting by the dyadic interaction (with the subject) high-rating group and the others.....	34
Figure 4.3	The results for the post-dialogue questionnaire regarding the dialogue evaluations in the triadic interaction setting by the triadic interaction high-rating group and the others.....	35
Figure 4.4	The results for the post-dialogue questionnaire regarding the impressions of the dialogue system in the dyadic interaction (with the subject) setting by the dyadic interaction (with the subject) high-rating group and the others.....	37

Figure 4.5	The results for the post-dialogue questionnaire regarding the impressions of the dialogue system in the triadic interaction setting by the triadic interaction high-rating group and the others.....	38
Figure 4.6	The interaction structure that the system constructed without the subject in the dyadic (with the topic) setting.....	39
Figure 4.7	The interaction structures desired by the subjects who were oriented toward dyadic interaction (blue dotted line) and the interaction structures in each system setting (red dotted line).....	40
Figure 4.8	The interaction structures desired by the subjects who were oriented toward triadic interaction (blue dotted line) and the interaction structures in each system setting (red dotted line).....	42
Figure 5.1	The results for the post-dialogue questionnaire regarding the dialogue system evaluations and the dialogue evaluations.....	46
Figure 5.2	Dendrograms from hierarchical cluster analysis for one of the subjects with the step number in brackets after each item.....	49
Figure 5.3	The distribution of the subjects for each number of steps from the dendrograms...	51
Figure 5.4	The distribution patterns.....	53

List of tables

Table 3.1	The prompts to ChatGPT and the processing of the dialogue system in the experiment.....	20
Table 5.1	Questionnaire responses for one of the subjects.....	47
Table 5.2	The classification of the system evaluation items.....	54

Chapter 1

Introduction

1.1 Dialogue Systems as Communication Partners

In recent years, interactive systems that humans can operate in natural language are in demand and have been put into practical use. Dialogue systems, such as customer support chatbots on corporate websites and virtual assistants, are becoming part of daily life. They are mainly used in temporary interactions replacing human workers. For example, answering vague questions, which is a task previously performed by humans in inquiry services, is handled through linguistic processing and searching. In the near future, along with the development of relevant technologies, dialogue systems are predicted to build longer-term and closer personal relationships with users [1]. Such relationships would be built to provide more personalized and effective support for users. If supportive relationships are built, mental benefits could be provided in addition to conventional practical benefits.

It has been shown that systems can demonstrate human-like value in communication. The media equation, which suggests that interactive systems can be perceived as equally human [2], is widely recognized. Conversations with chatbots have positive psychological effects similar to conversations with humans [3], and apologies from robots have value for people [4]. Through communication, people mentally influence each other, which directly affects their well-being and often triggers practical actions that affect their quality of life. As interactive systems could also influence users in a similar way, they are expected to provide support from both practical and mental aspects. Systems can be helpful partners for people supporting their life through communication in personal relationships. Dialogue systems equipped with advanced linguistic abilities are increasingly expected. This is because language plays a crucial role in human communication.

1.2 Human Communication Based on Language

Communication is essential for humans as social beings. Linguistic communication is especially important in social activities. Through conversation, people develop shared understandings and build relationships of trust and cooperation. Due to its importance, people have been interested in linguistic communication skills. There are many instructional books on speaking and listening skills [5, 6]. People's ultimate goals in using these skills are not merely to enhance linguistic ability. They generally aim to build relationships in which they are trusted and listened to by others through face-to-face conversation [7, 8]. Based on those relationships, they benefit from each other. Linguistic communication is a primary tool for building relationships.

Some people are engaged in linguistic communication to build trustworthy relationships in their profession. Healthcare professionals are a representative example. In the medical field, it is first necessary to build a relationship of trust in order to hear the patient's story, make a diagnosis, and provide treatment with the goal of achieving health. While treatment is conducted for the patient's benefit, this is not acceptable without trust. Therefore, methods for conducting interviews and engaging with patients are proposed [9, 10]. Particularly, mental health care providers give special attention to such relationships. Treatment is provided through linguistic communication within a trustworthy relationship [11, 12]. For such important benefits, interpersonal relationships built through linguistic communication are fundamental.

At the same time, human relationships can also have negative aspects. Relationships in care work are built for treatment based on their profession without involving personal selves. However, relationships tend to be personal. For example, personal self-disclosure is sometimes used during the process of relationship building in care work [13], and such relationships in their work involving personal feelings and stances can become a burden [14]. Not only in the medical field, but also jobs in the service industry involving conversation often carry a burden in relationships, so-called emotional labor [15]. These are negative effects resulting from unintended interpersonal relationships formed through situations that require linguistic communication.

It goes without saying that in personal relationships, people can be made happy or sad by what friends, family, and others say. People live in interpersonal relationships built through linguistic communication, giving, taking, and co-creating practical and mental benefits. Naturally, such relationships are a matter of great concern for them.

Research on interpersonal relationships has been conducted from multiple viewpoints.

Developmental science investigates how humans develop the ability to build relationships. Social developmental stages of recognizing and interacting with others as children become mature humans are studied [16]. In the fields of psychology and sociology, it is approached to resolve interpersonal conflicts and achieve well-being [17, 18]. The issues include how to tell others about oneself [19], how to build relationships at the first meeting [20], and what the ideal state of such relationships is, and how to maintain it [21]. Building better relationships is also addressed as an important issue in education. Students from kindergarten to university learn how to express their thoughts and emotions, how to convey and receive feelings with others in their society [22, 23, 24]. Social skills, the ability to build and maintain relationships with others effectively, have become highly valued in recent years [25]. These are a subject of considerable interest, including among adults. In these various fields, linguistic communication is regarded as an essential element in managing interpersonal relationships.

1.3 Dialogue Systems with Human-like Linguistic Capability

Generative AI using large language models has emerged, making dialogue with systems more human-like. Systems can generate new sentences as if they had their own mind. For example, ChatGPT-4 (OpenAI) demonstrated advanced linguistic abilities, answering a theory-of-mind test including a false belief scenario correctly, and the score matches the performance of six-year-old children [26]. Additionally, it was recognized as a human in a Turing test [27]. Dialogue systems based on generative AI have acquired advanced linguistic abilities that allow them to converse as if they were human.

To date, interactive robots and systems have tried to become partners with humans. They aimed to become friends [28], invited cooperation from humans [29], and talked about existing relationships involving users [30]. Even non-interactive products have been designed to evoke affectionate emotions in addition to their functions [31]. What they needed was to be recognized as members of human relationships, through attachment, collaboration, and relational knowledge, and to work as full participants in each situation. While they were welcomed, they have remained in a secondary role in most cases, such as in care work with pet robots [32]. They could not take part in conversations, although this is an important part of building relationships. Now, generative AI-based dialogue systems are emerging as conversational partners much like humans. They can

talk in natural language using broad knowledge and join conversations with human users. In human interpersonal relationships, linguistic communication is a primary tool for building relationships. Human-like linguistic capability would help interactive robots and systems become members of their users' interpersonal relationships.

However, generative AI-based dialogue systems have had mere ability to converse and have not had social ability to handle relationships yet. What matters to people is not conversation itself, but the interpersonal relationships built through it. Dialogue systems with human-like linguistic capability are expected to provide benefits similar to those in human relationships, that is, they need to handle relationships through conversation. On the other hand, although linguistic communication between humans has been widely explored, no single correct method has been identified for building and maintaining relationships. It is often concluded that people should manage relationships by adapting to each case and individual as fellow human beings. Therefore, what exactly systems should do has also been uncertain. It is necessary to explore how dialogue systems with human-like linguistic capability handle relationships through linguistic communication as social beings participating in human relationships.

Furthermore, when dialogue systems participate in human relationships with their social abilities, they can support not only human-computer interaction but also human-human interaction. In today's society, collaboration in confirming each other's intentions through linguistic communication has come to be valued in humans, rather than simply engaging in joint actions. Increasing attention to social skills is one reflection of this. Tools assisting with interactions will be welcomed, just like tools for strength or vision. Dialogue systems could stand by people as linguistic data processors, interpreters between people, and instructors in activities, sharing the effort in their personal relationships. They would be helpful partners for people supporting their life in relationships through linguistic communication from both practical and mental aspects.

Interpersonal relationships have various aspects, such as the handling changes depending on whether the relationship is being built or maintained. Dialogue systems will require much further development of social abilities to handle relationships. This thesis investigates how to build better relationships with humans as the first step of systems' social development, so that systems can become supporters of human well-being.

1.4 Structure of This Thesis

This thesis presents approaches for dialogue systems with human-like linguistic capability to become social and favorable partners in human life through communication. A human-like dialogue system based on generative AI was developed for investigation. The communication experiment between the system and human participants was conducted. The results of the system's operation and participants' subjective evaluations are analyzed and discussed. There are two concepts regarding the system: how it should interact as a social being, and how it can be personally accepted by individuals.

Chapter 1, Introduction, introduced the aim of this thesis with background.

Chapter 2, Human-like Dialogue Systems in Human-Computer Interaction, reviews previous research on dialogue systems, focusing on their development toward more human-like interaction. Social interaction ability and relationship compatibility are mentioned to indicate the two concepts of this thesis.

Chapter 3, Development of Generative AI-based Dialogue System, describes the development of the dialogue system and the design of the communication experiment. Outline of the dialogue system operation and the subjects' responses is indicated. Experimental results are summarized in this chapter and examined in detail from each viewpoint in the following chapters.

Chapter 4, Subjective Evaluation of Generative AI-based Dialogues, considers the construction of interaction structures. Subjective evaluations of the dialogues with the system in different interaction structure settings are investigated to clarify how the system interacts as a social being.

Chapter 5, Generative AI-based Dialogues Considering Individual Preferences, considers the properties of the dialogue system. Evaluation items are investigated to clarify what properties can be generally or personally well-received by individual subjects.

Chapter 6, Conclusions, summarizes the findings and describes future work.

Figure 1.1 shows the structure of this thesis.

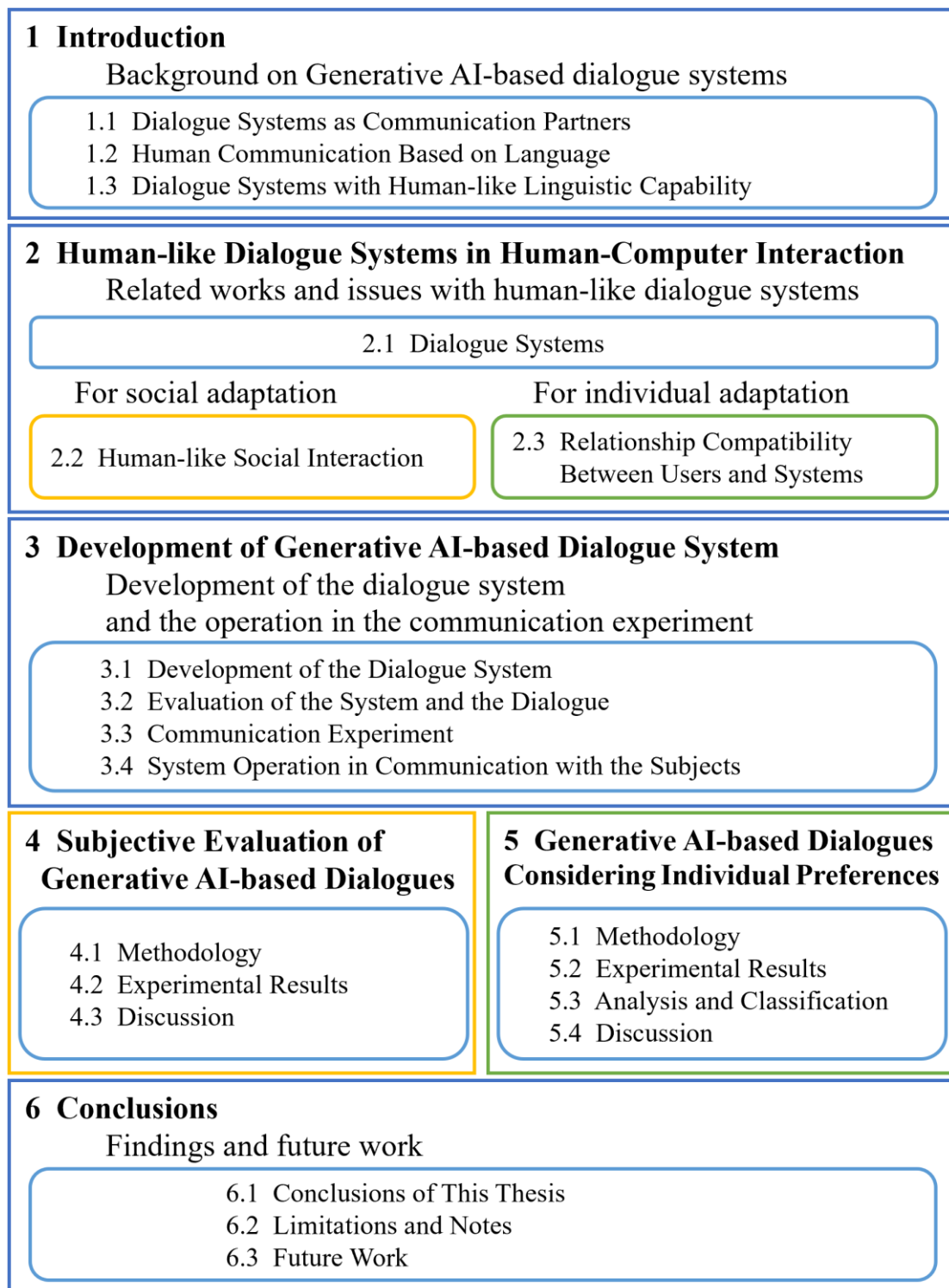


Figure 1.1. The structure of this thesis.

Chapter 2

Human-like Dialogue Systems in Human-Computer Interaction

2.1 Dialogue Systems

Dialogue systems are interactive systems that engage in communication with humans using language in the field of human-computer interaction. When dialogue is interactive, people often perceive their dialogue partner as human-like, even if it is clearly non-human. Even early dialogue systems using simple pattern matching gave a human-like impression. This is natural considering that similar conversational strategies are also used in human-human conversation, such as answering relevant words based on the conversation partner's utterance and repeating what the partner said [33]. Assisted by such human recognition, interactions with dialogue systems have been accepted as conversations without particular difficulty. Conversational human-computer interaction with dialogue systems is helpful due to its ease of use. Users can operate systems in a way similar to everyday human-human conversations, without prior knowledge of the systems. Dialogue systems with easy to talk to appearances have shown a certain level of success in actual use, such as guidance characters on digital signage installed at train stations [34].

More human-like conversations have been considered effective in improving relationships between users and systems. Some conversational software applications have been developed to make interactions more human-like by introducing emotions, personas, or personal stories [35, 36, 37]. Contests involving systems that engage in casual conversation with people have been held [38]. These systems have been greeted with joy and attention. However, it has also been shown that human-like performance is sometimes perceived as unpleasant [39]. The uncanny valley hypothesis, which concerns eerie feelings arising as human-likeness increases, is inconclusive [40]. This is still a complex issue regarding user feelings toward human-like systems,

including dialogue systems. Previous research on chatbots' responses shows some necessary characteristics for positive user experiences, while the answers to each question, choosing the preferred chatbot's response, in the survey, were mixed [41]. The best response differs from person to person.

Dialogue systems have developed and become more human-like. As a result, they have come to need to adapt their behavior to users' situations and personal preferences. In personal relationships, it is natural that people have various ways of interacting with others and perceive them differently, depending on personal factors. When systems relate to users personally, they are required to adapt to the range of individual differences. The more personal the relationship, the more important it is that the behavior during interactions is acceptable to individual users. Behaviors of dialogue systems are further developing. Systems have acquired advanced linguistic abilities using generative AI, have developed their non-verbal behavior in conversation [42], and are attempting to handle participants' emotions in conversations [43]. Such advanced features would improve the quality of conversations with systems, and advanced behavior resulting from those would make users more selective. Dialogue systems that communicate with human-like capability should develop their behavior while adapting to each user's purpose and preference.

2.2 Human-like Social Interaction

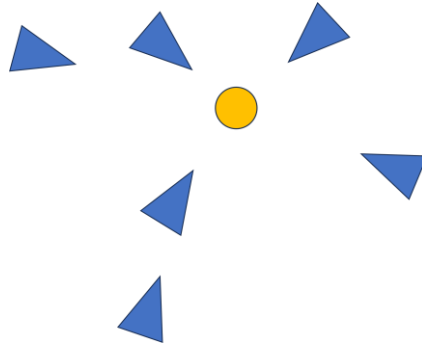
As stated in Chapter 1, linguistic communication is essential in human society. Because of the importance, generative AI-based dialogue systems that can converse like humans attract high expectations. Language is used to influence human behavior. Next-generation AI agents [44] including management of everyday issues in healthcare, education, and whole lifestyles [45, 46, 47] need dialogue systems as actuators to influence and assist users' activities using their linguistic capability. For such an interventional purpose, it is necessary to consider that humans are unlikely to accept words without basic relationships. They first build cooperative relationships through conversations to confirm their situation and reliability before accepting intervention. If systems work using language in order to assist and sometimes instruct human behavior, they should first build supportive relationships with users as well. Building relationships is precisely a social action. They should interact with users socially in human-like ways, adapting to social contexts according to their conversational partners' intentions and expectations. Systems' linguistic abilities need to

be used with social abilities.

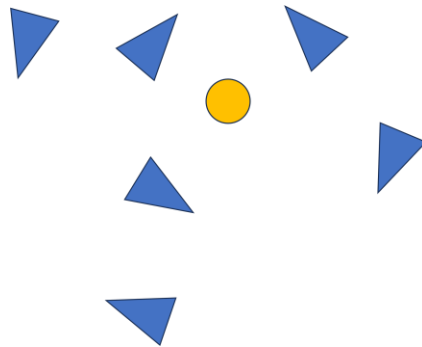
The importance of social interactions with users is shown in other perspectives. While humans tend to perceive only living beings as communication partners, they sometimes feel a sense of animacy, recognizing non-living things as living [48]. Especially, when the objects indicate social or communal relationships interacting with each other, a sense of animacy and empathy toward the objects can arise even if the objects are geometric figures [49, 50]. An example of animacy on geometric figures is shown in Figure 2.1. Additionally, although ordinarily brain cortical activity differs when interacting with objects versus living beings [51], it has been shown that the brains of patients with autism and Asperger syndrome who have social communication difficulties indicate activity similar to that observed in object interactions, whether they are interacting with humans or objects [52]. Communication and building relationships through it are based on recognizing someone or something as a partner capable of social interaction. Interacting socially would help dialogue systems be recognized as partners.

Regarding human social interaction ability, it has been shown that infants have higher level interaction abilities for social relationships, such as sharing goals, compared to apes [16]. These abilities develop through interactions with others. Children learn how to relate to people by imitating familiar adults' behavior in interactions [53]. The ability to build relationships gradually develops through interactive communication. Some behaviors observed in early childhood that attempt to share attention with a partner are shown in Figure 2.2. Since human attention is shown in head movement [54], turning one's face toward something of interest and toward what a partner is gazing at is a way of sharing interests in social communication. Children begin to visually follow people and objects around them as their eyesight develops from around three months, and they also start to make eye contact. Next, they attempt to share attention with others. This is called joint attention in the field of developmental science, an essential development for social relationships in human communication at around nine months [16].

Development of social abilities in children, such as joint attention, has been referenced and applied to communication robots to improve their interaction in the field of developmental robotics [55, 56]. For instance, the fundamental communication skills related to theory of mind are applied to robots and show the potential for application [57]. Therefore, applying human-like social interaction behaviors in conversation to dialogue systems is expected to help build relationships with humans. Particularly, the essential social behaviors, which are observed during the developmental stages of children, would be necessary for systems.

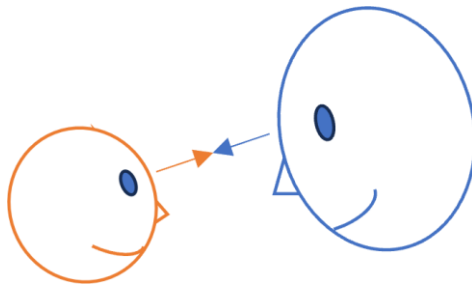


(a) The acute angles of the triangles are pointing toward the circle.

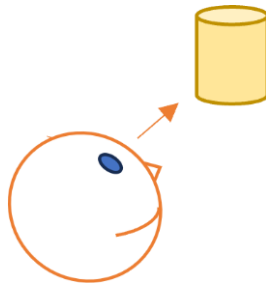


(b) The acute angles of the triangles are oriented at 90 degrees to the direction of the circle.

Figure 2.1. An example of animacy. The positions of the triangles and the circle are the same in both (a) and (b). When people move the circle, they avoid the triangles feeling targeted even if it is not necessary, more in the (a) case than in the (b) case [49].



Eye contact.



Find an object.



Joint attention. Direct the partner's attention by pointing, and check the attention.

Figure 2.2. Examples of behaviors in early childhood. Children recognize the presence of others and objects, and then attempt to share the things around them.

Therefore, the dialogue system for the communication experiment was designed to be equipped with human-like social interaction behavior, with reference to child development. Chapter 3 describes the dialogue system and the communication experiment. Chapter 4 discusses the system behavior focusing on interaction structures that communication participants construct during conversation. The system participated in the constructions as a social being in the interactions with the human subjects.

2.3 Relationship Compatibility Between Users and Systems

As is often remarked, relationship compatibility is a key factor in personal relationships. Advanced human-like dialogue systems would be liked or disliked depending on users' varied preferences of interactions in interpersonal relationships. For systems, personality adaptiveness can become a valuable feature [58, 59]. To be accepted in human personal relationships, systems should adjust their properties depending on users' preferences and the roles expected of them. However, adjustments of system properties often include contradictory factors. For example, politeness is one of the standard characteristics of dialogue systems [41]. In general, polite people tend to be more liked, and systems are similarly liked. On the other hand, people sometimes behave impolitely in close relationships to express their familiarity. In practice, while impolite behavior of systems was evaluated by users as having humanity, it caused unintended discomfort involving a risk of negative evaluations [60].

As an approach, system properties can be designed in correspondence to personality traits in humans. It is possible to use characters whose personality traits have already been established. This approach can be effective for users who prefer personality-driven system behaviors [61]. Positive feedback can be received through consistent behaviors based on the established personality. Nevertheless, being anonymous or non-human is one of the distinctive features of dialogue systems. This is sometimes an advantage in communication with users [62, 63]. The behavior expected of systems differs from the behavior expected of humans. Users behave differently toward humans and chatbots [64, 65], and productivity is the primary expectation for chatbots [66]. Desired properties of dialogue systems with human-like linguistic capability may differ from preferred personality traits in humans or characters.

Therefore, it is necessary to clarify what properties of dialogue systems influence users'

evaluations. Additionally, system properties would have different effects depending on individual preferences. The differences should be considered to adapt to individual users. Chapter 5 investigates the system properties considering individual preferences in system behavior to improve relationship compatibility between users and systems for more favorable relationships.

Chapter 3

Development of Generative AI-based Dialogue System

The dialogue system was developed using generative AI for the communication experiment, which was designed for one-on-one linguistic communication with the system as a partner. Human-like conversations were achieved by generating response texts in real time. The system communicated with the subjects in natural language via voice and featured a facial expression function, reflecting the fact that people communicate with others in voice-based and face-to-face interactions in everyday situations. To investigate different types of behavior in communication, the settings of dyadic and triadic interactions were applied to the system. These settings referenced the change of interaction structures from dyadic to triadic in the developmental stages of children regarding social communication. The dialogue system and the dialogue with the system were evaluated through a questionnaire. The system demonstrated its communication capability in the experiment.

3.1 Development of the Dialogue System

3.1.1 System Overview

A laptop PC was used for the dialogue system. The subjects input their voice into the PC microphone. The dialogue system output voice, reading texts, from the PC speakers and displayed facial expression images on the screen. The PC screen was masked except for the facial expressions. An overview of the dialogue system is shown in Figure 3.1.

The dialogue was conducted in a question-and-answer format. The subject switched on the microphone to input a voice comment and then switched it off. The dialogue system responded to this. The system did not refer to what the subject had said in the previous turn. A generative

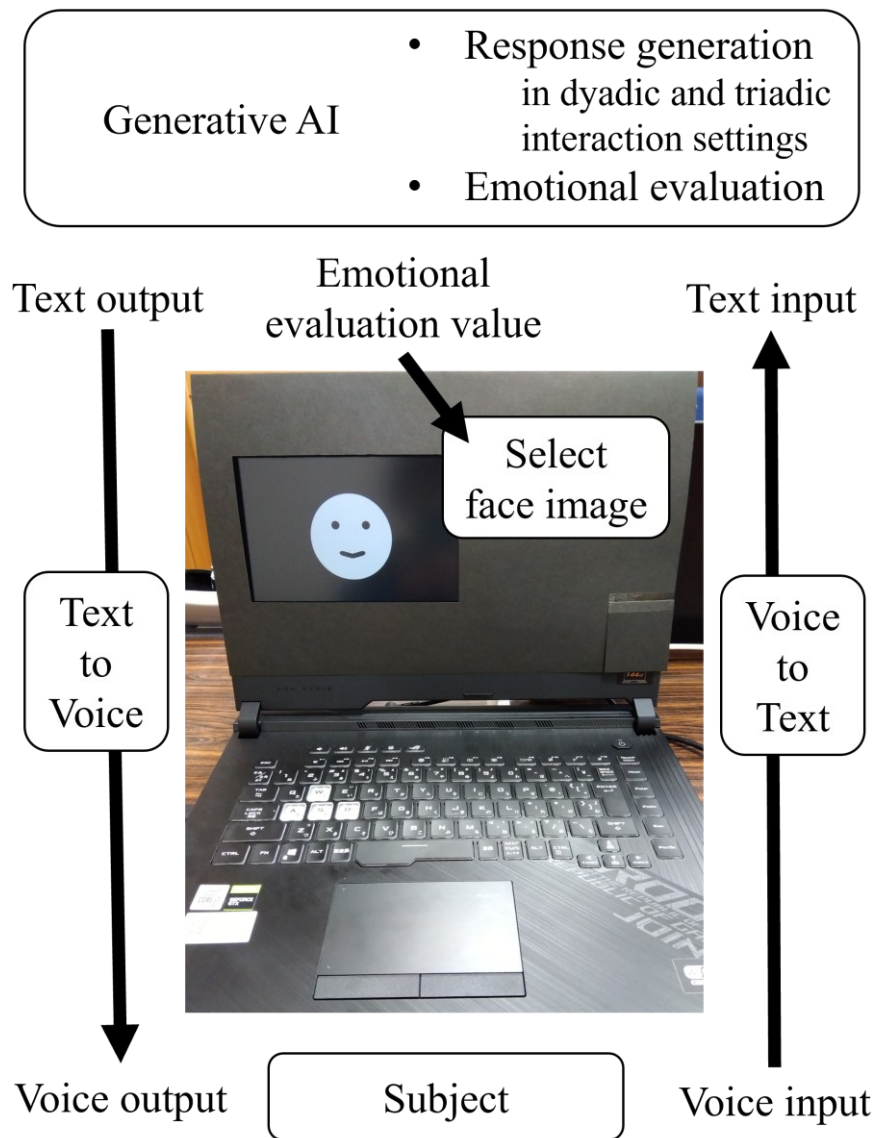


Figure 3.1. An overview of the dialogue system.



Figure 3.2. The face images displayed on the screen for five levels of emotional expression.

AI, ChatGPT (gpt-3.5-turbo-0613 model, OpenAI), was used to generate sentences for the dialogue system's voice output.

The dialogue system for the communication experiment was a human-like communication partner. To help the subjects feel this, a facial expression function was developed. It is known that emotional information can be conveyed to humans through expressions by machine systems. Regarding the emotional expressions, the effectiveness of combining different forms of expression [67], using human-like expressions [68], and using color-based representations [69] has been shown. Accordingly, the face image displayed on the screen of the dialogue system changed the shape of the mouth and the face color, as shown in Figure 3.2. There were five levels of emotional expression, with the images on the left side used to express more negative emotions and the images on the right side used to express more positive emotions. To select the face image related to a dialogue, ChatGPT (gpt-3.5-turbo-0613 model, OpenAI) was used. This feature did not classify emotions in detail but simply determined whether the emotion was positive or negative. It is known that cultural differences exist in emotional expression and classification between Eastern and Western cultures. However, “happy” and “sad” are commonly used for positive and negative emotions [70]. Therefore, the prompts for emotional evaluation included the degree of happiness and sadness.

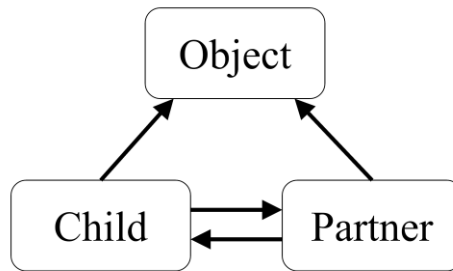


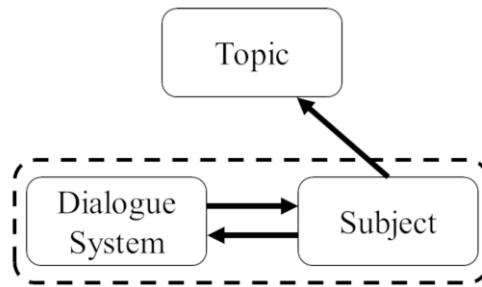
Figure 3.3. Triadic interaction structure between a child, a partner, and an object of interest.

3.1.2 Settings of Dyadic and Triadic Interactions

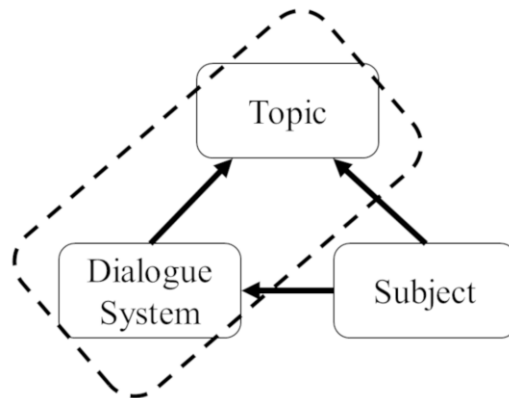
The dialogues were conducted in three types of interaction settings with reference to the development of social interaction ability in early childhood. The referenced development is the emergence of joint attention around nine months, which is considered an essential development for social relationships in human communication [16, 71]. During stages of communication development in children, relationships with someone or something begin as a one-to-one dyadic interaction between oneself and a partner or between oneself and an object of interest. Next, relationships progress to a triadic interaction involving joint attention between oneself, a partner, and an object of common interest, as shown in Figure 3.3. At this stage of development, children share objects with others and begin to cooperate as social beings.

This development of interaction structures from dyadic to triadic was applied to the interaction structure between the dialogue system and the subject in the system settings. The dialogue system interacted in the role of the child shown in Figure 3.3. The subjects were assigned to the partner role. The topics were assigned to the object of common interest, since talking about a shared topic was considered a joint and cooperative action. Figure 3.4 shows the interaction structures between the subject, dialogue system, and topic. The system operated to construct the structures enclosed by the dotted line in Figure 3.4 according to each setting.

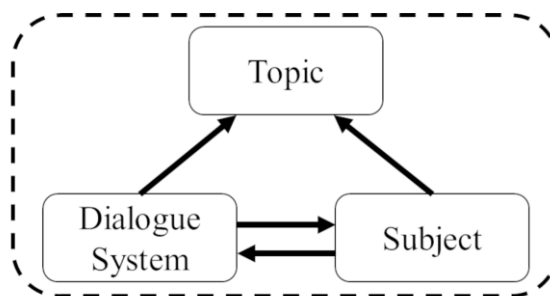
Taking the case where the topic is “Which is better, orange or apple”, as an example, the



Dyadic interaction
(dialogue system with subject)



Dyadic interaction
(dialogue system with topic)



Triadic interaction

Figure 3.4. The interaction structures between the subject, the dialogue system, and the topic, according to the settings of the dialogue system.

behaviors of the system in each setting were as follows:

- In the dyadic interaction (between the system and the subject) setting, the system responds to what the subject says. If the subject goes off-topic, the system follows the subject. For example, the subject asks, “What do you like?”, the system responds, “One of my absolute favorite hobbies is traveling. Exploring new places and experiencing different cultures brings me so much joy!”
- In the dyadic interaction (between the system and the topic) setting, the system reads out the prepared sentences one at a time and does not respond to what the subject says. For example, the subject asks, “What do you like?” the system responds, “Both oranges and apples have their own wonderful qualities!”
- In the triadic interaction (between the system, the subject, and the topic) setting, the system responds on topic to what the subject says. If the subject goes off-topic, the system continues to talk about the topic and sometimes points it out. For example, if a subject asks, “What do you like?” the system responds, “I enjoy both oranges and apples, but my choice depends on my mood or the season. They’re both nutritious and delicious fruits, don’t you think?”

3.1.3 Algorithms

The prompts and the processing for the dialogue system in each setting are shown in Table 3.1. In the vertical axis of Table 3.1, the actions of the subject and system are listed with the progression of the experiment. In the horizontal axis, three types of interaction settings are arranged. The subjects and the system communicated in Japanese, and the prompts were also written in Japanese.

To generate responses, ChatGPT (gpt-3.5-turbo-0613 model, OpenAI) was prompted in its system role to respond to a subject input in a polite tone with approximately 25 words, and in its assistant role to provide a reply to a subject input or to provide a reply to a subject input on a specified topic.

To evaluate emotions, ChatGPT (gpt-3.5-turbo-0613 model, OpenAI) was prompted in its assistant role to provide a happiness degree and a sadness degree on a scale of 100, evaluating the emotions present in a combination of a subject input and a generated response to this. The difference in numbers between the degree of happiness and sadness was converted into five levels with equal intervals, and these levels were used to select a face image.

Unity (Unity 2022.3.0, Unity Technologies) was used to integrate and present these functions.

Table 3.1. The prompts to ChatGPT and the processing of the dialogue system in the experiment.

Dialogue progression	Interaction setting		
	Dyadic interaction (with subject)	Dyadic interaction (with topic)	Triadic interaction
Start	Neutral level facial expression image is displayed, and a standard phrase is announced, e.g., “Hello. The first topic is which fruit is better, orange or apple. Let’s talk.”		
Subject turn	Some voice input, e.g., “I like an orange. I think, an orange is better.”		
Processing	Prompt for a response: Respond to a subject input in a polite tone with approximately 25 words.	All response texts and face image selections are prepared in advance.	Prompt for a response: Respond on topic to a subject input in a polite tone of approximately 25 words on a topic.
	Prompt for emotional evaluation: Rate happiness and sadness on a scale of 100 about “a subject input and a generated response”. The difference in numbers between the degree of happiness and sadness is changed to five levels and used to select a face image.		Prompt for emotional evaluation: Rate happiness and sadness on a scale of 100 about “a subject input and a generated response”. The difference in numbers between the degree of happiness and sadness is changed to five levels and used to select a face image.
System turn	A selected face image is displayed, and a response text is read out.		
Repeat Subject turn, Processing, and System turn for 5 min.			
End	A standard phrase is announced, “Thank you. That’s all for this topic. Please fill out the questionnaire. Please talk to me after you’ve finished.”		

3.2 Evaluation of the System and the Dialogue

Japanese questionnaires were created for subjective evaluations: the pre-experiment questionnaire and the post-dialogue questionnaire. The subjects were asked to respond to each item on a six-point visual analog scale.

Before the experiment, the subjects' conditions were surveyed. If the subjects' mood, sociability, and expectations for dialogue systems were especially negative, the evaluations would be affected. The pre-experiment questionnaire confirmed these.

The subjects were asked about themselves using the following items: "Your mood" from negative to positive, "Meeting people", "Listening to people", "Speaking to people", and "Becoming friends with people" from dislike to like, and "Talking to AI" (To make it easier for the subjects to understand, the dialogue system was referred to as AI.) from negative to positive.

Additionally, a free-form comment section was provided. The pre-experiment questionnaire sheet is shown in Appendices.

The post-dialogue questionnaire was provided for evaluations of the system and the dialogues.

For the dialogue system evaluation, items were developed with reference to research on communication partners in Japan. The items from the research on the attractiveness factors of a friend [72] were as follows: "Independent", "Sincere", "Inspiring", and "Easy to talk to". The items from the research on the impression of the communication robot [73] were as follows: "Awkward", "Has personality", and "Inhuman". Additionally, "Friendly" was included with the expectation that a friendly system would be evaluated positively as a dialogue partner.

For the dialogue evaluation, some subjects were asked what types of dialogue they consider to be good in advance. Dialogue evaluation items were developed based on the answers. The items for the dialogue evaluation were as follows: "Fun", "It listens", "Can have a smooth conversation", and "Want to talk about other topics too".

Additionally, "Your mood" from negative to positive was asked, and a free-form comment section was provided. The post-dialogue questionnaire sheet is shown in Appendices.

3.3 Communication Experiment

3.3.1 Conditions

The dialogue system conducted first-meeting conversations with the subjects in the experiment. The subjects were only told that there were three types of partners and were not given any prior information about their respective settings. Social abilities for building relationships in first meetings are exhibited in order to be recognized as a favorable conversational partner. In a first meeting scene of interpersonal relationships, after talking for a while, one judges whether the other person is someone who can communicate well with them, based on their satisfaction with the conversation. The system interacted with the subjects in three interaction settings with different stages of social ability development. The subjective evaluations of the system and the dialogues were examined to investigate the effect of these differences on satisfaction with the dialogues.

To avoid interference from others in the relationships between the system and each subject, the experiment was conducted in a room with only the subject, and the contents of the dialogues were not saved. The time series of the number of characters spoken during the experiment was saved and used to verify that a dialogue had taken place.

3.3.2 Topics

The topics in the experiment were “Which is better: orange or apple, airplane or bullet train, and summer or winter”, which did not contain a strong meaning or emotion for the subjects. When subjects were asked to talk about emotional topics in pilot experiments, the results showed differences in the degree of feelings depending on their chosen topics, and some subjects felt a resistance to talking. Although exchanging emotions is important in communication, research on emotional conversations and conversation contents was not the purpose of this study. Therefore, topics with a strong meaning or emotion were avoided to ensure the equality of the experimental conditions.

3.3.3 Subjects

The subjects were 13 healthy adults (5 males, 8 females, age 35.8 ± 14.4 years) including university students and working adults. It was confirmed in advance that they do not feel specific difficulty with communication.

The contents of the experiment were explained to the subjects in advance. Informed consent was obtained from all subjects involved in this research. The experiment involving human participants was conducted with careful consideration in accordance with Research Ethics at Kobe University.

3.3.4 Experimental Procedure

1. The subject inputs any voice, such as hello, into the system.
2. The dialogue system announces a topic and starts a dialogue.
3. After five minutes, the system announces the end of a dialogue and asks the subject to fill out the post-dialogue questionnaire.

These steps were repeated for the three types of system settings. When three dialogues were completed, the dialogue system announced the end of the experiment.

When the experimenter explained how to talk with the dialogue system, the subjects practiced talking with the system for up to two turns. Considering that impressions would change depending on order, the order of the settings was switched for each subject to ensure equality. The six possible orders for the three settings were assigned to the subjects in the order of their participation in the experiment. Regardless of these orders, the topics were presented in the following fixed order: fruits, transportation, and seasons.

The experimental procedure is shown in Figure 3.5 and an overview of the system workflow in the experiment is shown in Figure 3.6.

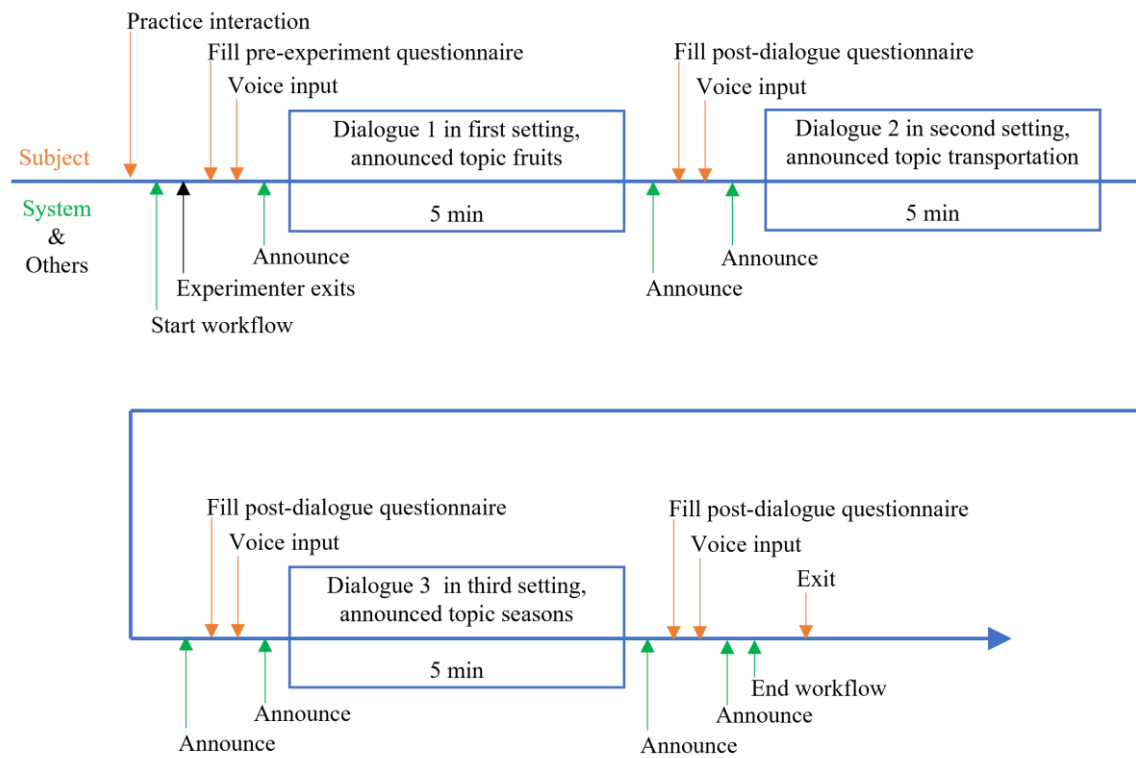


Figure 3.5. The experimental procedure.

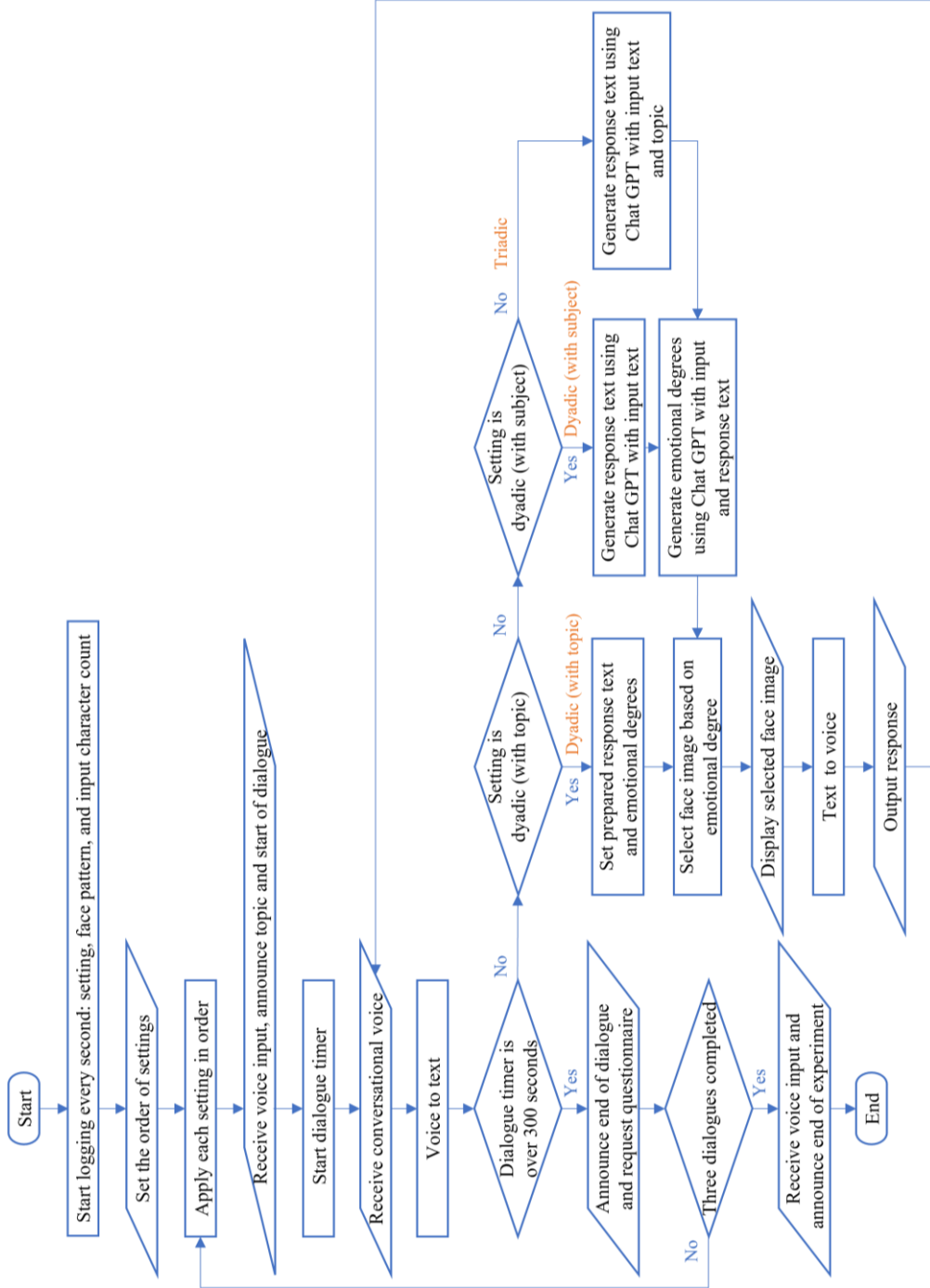


Figure 3.6. An overview of the system workflow.

3.4 System Operation in Communication with the Subjects

All the subjects who participated in the experiment communicated with the dialogue system. They responded to the pre-experiment questionnaire and evaluated the system and the dialogues through the post-dialogue questionnaire.

The average and standard deviation of the results for the pre-experiment questionnaire are shown in Figure 3.7. In the responses, one subject rated “Talking to AI” item the lowest value of 1. This was the only one especially negative response. All the other ratings were between 3 and 6. The one subject who dislikes dialogue with systems evaluated the dialogue in the dyadic interaction (with the topic) setting the highest. The subject might have perceived the one-sided system as the system requiring the least effort to communicate.

The average and standard deviation of the results for the post-dialogue questionnaire are shown in Figure 3.8. The system behaved differently in each applied setting. The system in the dyadic interaction (with the topic) setting, ignored the subjects in their interactions. The dialogue evaluations in this setting were negative, and the dialogue impossibility was indicated in the free-form comments. In the dyadic interaction (with the subject) and triadic interaction setting, the dialogues took place between the subjects and the system. The evaluations were positive on average, and there was no mention of the impossibility of dialogue in the comments.

The evaluation of dialogue quality was examined based on the total evaluation value for the four evaluation items about dialogue. There was a strong positive correlation ($r=0.9$) between this total value and the value of the “Your mood” item in the post-dialogue questionnaire, as shown in Figure 3.9. The total value reflected the satisfaction with the dialogue.

The system’s face with emotional expressions was not mentioned in the free-form comment section by any subject. The facial expression changed 18.0 ± 4.4 times during the three dialogues for each subject in the experiment, and the number of displays for each expression level was as follows, from the most negative to the most positive: 0 , 2.5 ± 1.7 , 4.2 ± 1.9 , 7.1 ± 2.1 , and 4.2 ± 1.4 times. This feature was naturally accepted by the subjects and did not leave a strong impression. It seems that the system naturally expressed appropriate emotional responses to the subjects. However, the experimental condition avoided emotional topics. Considering that the most negative expression was never used, the dialogues in the experiment were likely light and polite, typical of a first meeting. In such dialogues, the system’s somewhat positive emotional responses probably did not have meaningful impacts. Further research on emotional dialogues

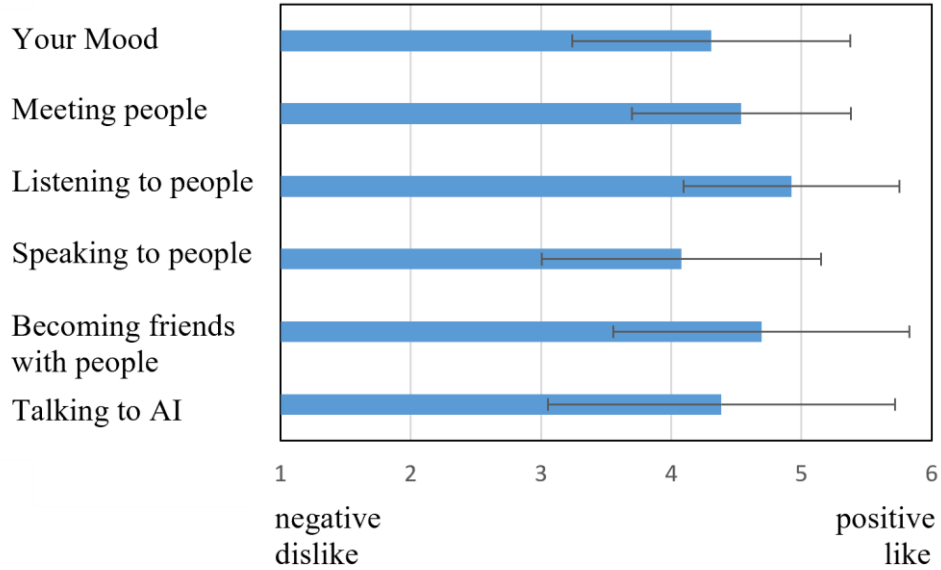


Figure 3.7. The results for the pre-experiment questionnaire.

could provide additional findings regarding the effects of systems' emotional expressions.

The system became a conversational partner in natural language in the two settings. The subjects variously evaluated the system and the dialogues. When the dialogue system repeated what they said, some subjects evaluated it as reassuring that the system understood what they were saying, while the others evaluated it as mechanical and parroting. These were contrasting evaluations for the same behavior of the system. The results of evaluations are investigated in detail in the following chapters. In Chapter 4, desired behaviors of the system are discussed from perspective of the social interaction. In Chapter 5, the properties of the system are classified based on the cluster analysis results of the links in the evaluations focusing on individual preferences.

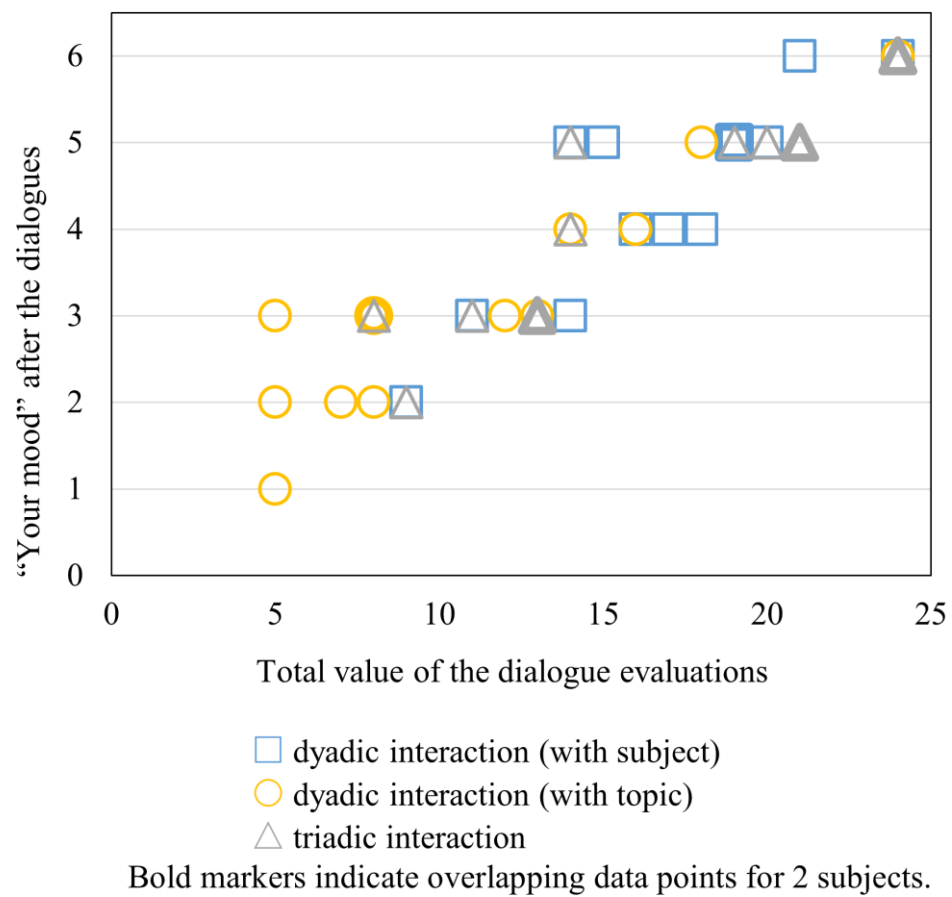


Figure 3.9. The relationship between the dialogue evaluations and the subjects' mood after the dialogues.

Chapter 4

Subjective Evaluation of Generative AI-based Dialogues

In this chapter, subjective evaluations of dialogues are investigated in the settings of paired dyadic and topic-sharing triadic interaction structures. The settings referenced the change of interaction structures from dyadic to triadic in the developmental stages of children regarding social communication.

The communication experiment was conducted under the hypothesis that the progressive development of interaction structure in the system settings would improve the system's interaction and provide favorable dialogues for the subjects. The hypothesis was partially confirmed. The dialogues with the system received positive evaluations on the basis of well-constructed interaction structures between the subjects and the system. When the construction failed, the evaluations of the dialogue and the system worsened. If the construction succeeded, the dialogues received more positive evaluations in the setting of the triadic structure, which is a developed form of interaction. The structure in the system setting needed to match the structure that the subject was oriented toward.

4.1 Methodology

Dialogue systems with human-like linguistic capability should behave socially for building relationships with humans. Essential social behaviors are observed during the developmental stages of children. Therefore, the dialogue system for the communication experiment was designed to be equipped with human-like social interaction behavior referencing child development. The referenced development was the change of interaction structures from dyadic to triadic.

The system started the dialogues in the experiment with a standard phrase that invited the

subjects to talk about the topic, e.g., “Hello. The first topic is which fruit is better, orange or apple. Let’s talk.” After such a phrase, the system played a designed role in the dyadic and triadic structures depending on the interaction settings. In the dyadic (with the subject) setting, it responded to what the subject said. In the dyadic (with the topic) setting, the system talked about only the topic. In the triadic setting, it responded on topic to what the subject said.

The triadic interaction structure is an advanced structure emerging in a later developmental stage than the dyadic one. The system could cooperate by sharing the topic in the triadic setting, in other words, it could engage in a discussion on the topic. The hypothesis was that the triadic interaction setting would improve the system’s behavior and provide favorable dialogues for the subjects. However, some subjects’ evaluations indicated that the dialogues with the system in this advanced setting were not comfortable. How the system should participate in the construction of interaction structures as a social being is considered based on the results.

4.2 Experimental Results

All the subjects interacted with the system and evaluated the dialogues and the system through the post-dialogue questionnaire. First, overall trends in the dialogue evaluations are presented. Second, the dialogue evaluations in the dyadic and triadic settings are described according to the subject groups divided based on their high-rating setting. Third, the system evaluations regarding the impressions it made on the subjects are presented.

4.2.1 Dialogue Evaluation

The average and standard deviation of the results for the dialogue evaluation items in the post-dialogue questionnaire are shown in Figure 4.1. In the remainder of this chapter, the items in the post-dialogue questionnaire with a mean evaluation value of 3.0 or less are treated as negated items, and items with a mean of 4.0 or more as affirmed items. Comments in the free-form section are cited in relevant discussions. In the results of the dyadic interaction (with the topic) setting, all four items were negated. These were negative results compared with the evaluations in the other two settings. The results of the dyadic (with the subject) and triadic interaction settings showed that “Fun”, “It listens”, and “Want to talk about other topics too” were affirmed in

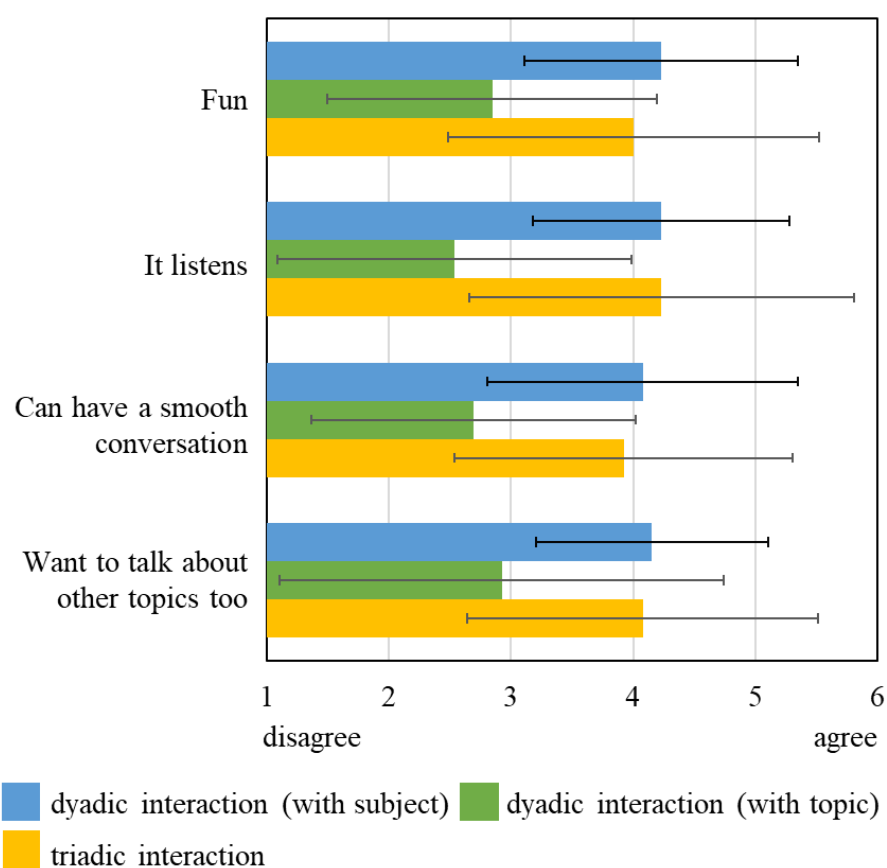


Figure 4.1. The results for the post-dialogue questionnaire regarding the dialogue evaluations.

common. However, the results in the triadic interaction setting showed many differences across the subjects. The following section exams this in more detail.

4.2.2 Dialogue Evaluation from High-Rating Groups in the Dyadic and Triadic Settings

The subjects were divided into three groups based on the sum of evaluation values for the four items regarding the dialogue in each interaction setting. The dyadic interaction (with the subject) high-rating group consisted of 5 subjects, and their evaluation values were as follows: dyadic (subject) 18.2 ± 2.5 , dyadic (topic) 9.6 ± 5.1 , and triadic 11.0 ± 2.3 . The dyadic interaction (with the topic) high-rating group consisted of 1 subject, and their evaluation values were as follows: dyadic (subject) 9, dyadic (topic) 16, and triadic 13. The triadic interaction high-rating group consisted of 6 subjects, and their evaluation values were as follows: dyadic (subject) 15.5 ± 2.6 , dyadic (topic) 9.2 ± 3.0 , and triadic 19.8 ± 3.0 . There was one subject who gave the same rating of 24 in all three settings. Since this result is unsuitable for comparison, it is excluded from the following discussion. The dyadic interaction (with the topic) high-rating group was small with one subject, and the comparison of the results focusing on this group is omitted.

Figure 4.2 shows the averages and standard deviations of the results in the dyadic interaction (with the subject) setting by the dyadic interaction (with the subject) high-rating group and the others. The Wilcoxon rank-sum test was conducted to compare the dyadic interaction (with the subject) high-rating group and the others with the evaluation value of the dialogue. The results indicated no significant difference between the two groups ($p=0.088$). Comparing the average values of the results between the two groups, there were no notable differences such as reversing the affirmed and negated evaluations for the same item.

Figure 4.3 shows the averages and standard deviations of the results in the triadic interaction setting by the triadic interaction high-rating group and the others. The Wilcoxon rank-sum test was conducted to compare the triadic interaction high-rating group and the others with the evaluation value of the dialogue. The results indicated a significant difference between the two groups ($p=0.005$). The subjects in the high-rating group evaluated the items “Fun”, “It listens”, “Can have a smooth conversation”, and “Want to talk about other topics too” as affirmed. The other subjects evaluated the items “Fun”, “It listens”, and “Can have a smooth conversation” as negated.

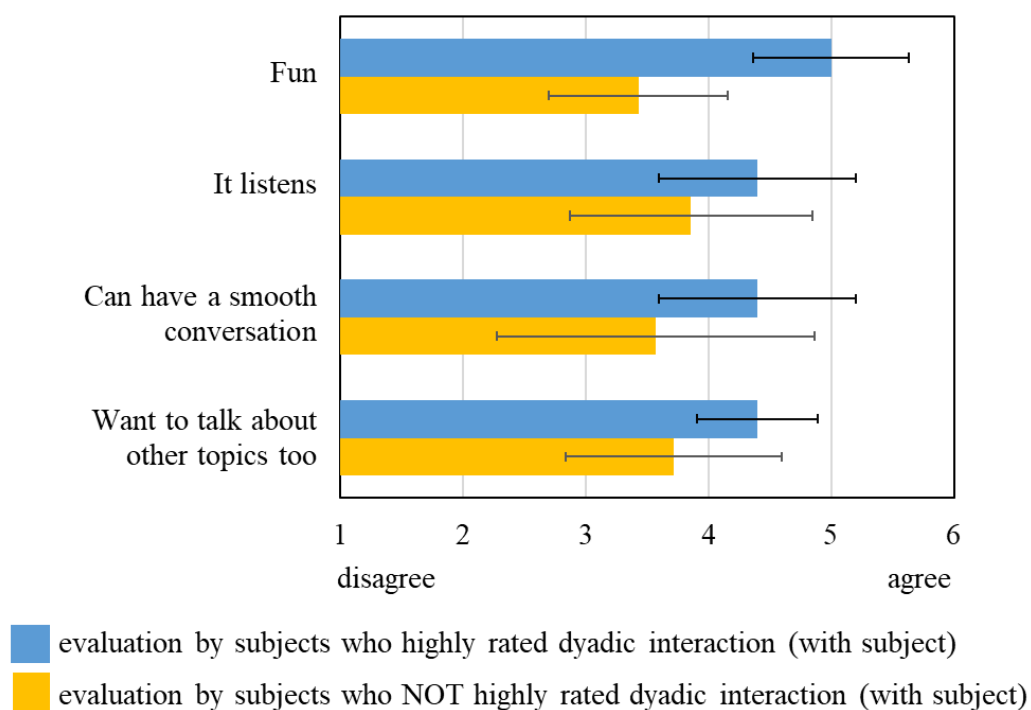


Figure 4.2. The results for the post-dialogue questionnaire regarding the dialogue evaluations in the dyadic interaction (with the subject) setting by the dyadic interaction (with the subject) high-rating group and the others.

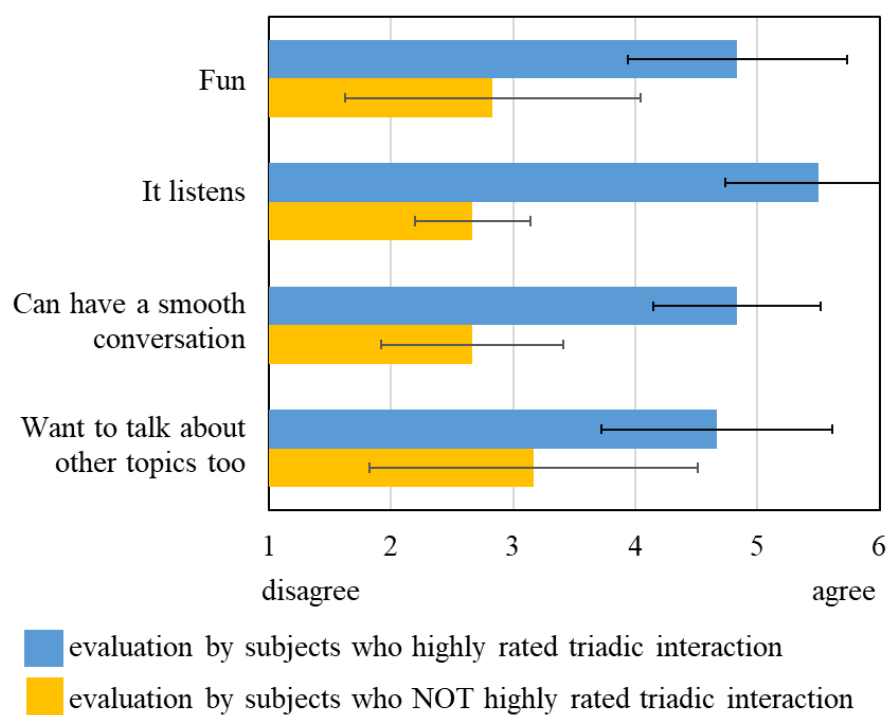


Figure 4.3. The results for the post-dialogue questionnaire regarding the dialogue evaluations in the triadic interaction setting by the triadic interaction high-rating group and the others.

4.2.3 Dialogue System Evaluation

Figure 4.4 shows the averages and standard deviations of the results for the items about impressions of the dialogue system in the dyadic interaction (with the subject) setting by the dyadic interaction (with the subject) high-rating group and the others. Comparing the average values of the results in the dyadic interaction (with the subject) setting, there were no notable differences such as reversing the affirmed and negated evaluations for the same item. The subjects commonly evaluated the items “Friendly”, “Independent”, and “Sincere” as affirmed and “Awkward” and “Inhuman” as negated.

Figure 4.5 shows the averages and standard deviations of the results for the items about impressions of the dialogue system in the triadic interaction setting by the triadic interaction high-rating group and the others. The subjects in the high-rating group evaluated the items “Friendly”, “Independent”, “Sincere”, “Inspiring”, “Has personality”, and “Easy to talk to” as affirmed and “Awkward” and “Inhuman” as negated. The others evaluated the items “Awkward” and “Inhuman” as affirmed and “Inspiring”, “Has personality”, and “Easy to talk to” as negated. The evaluations were reversed for five items.

4.3 Discussion

In the experiment, the subjects communicated with the human-like dialogue system using generative AI in the three types of interaction settings and evaluated the dialogues and the system. The settings of interaction structures changed the system behavior. The dialogue evaluations in the dyadic interaction (with the topic) setting were negative, and the dialogue impossibility was indicated in the free-form comments. In the dyadic interaction (with the subject) setting and triadic interaction setting, the dialogues took place between the subjects and the system. The evaluations were positive on average, and there was no mention of the impossibility of dialogue in the comments.

The triadic interaction setting was applied to enable more socially advanced interaction. The evaluations were expected to be higher than those in the dyadic settings. The results partially confirmed this expectation. The evaluations of the dialogue in the triadic interaction setting were both highly positive and negative, and the impressions of the system were contrasting among the

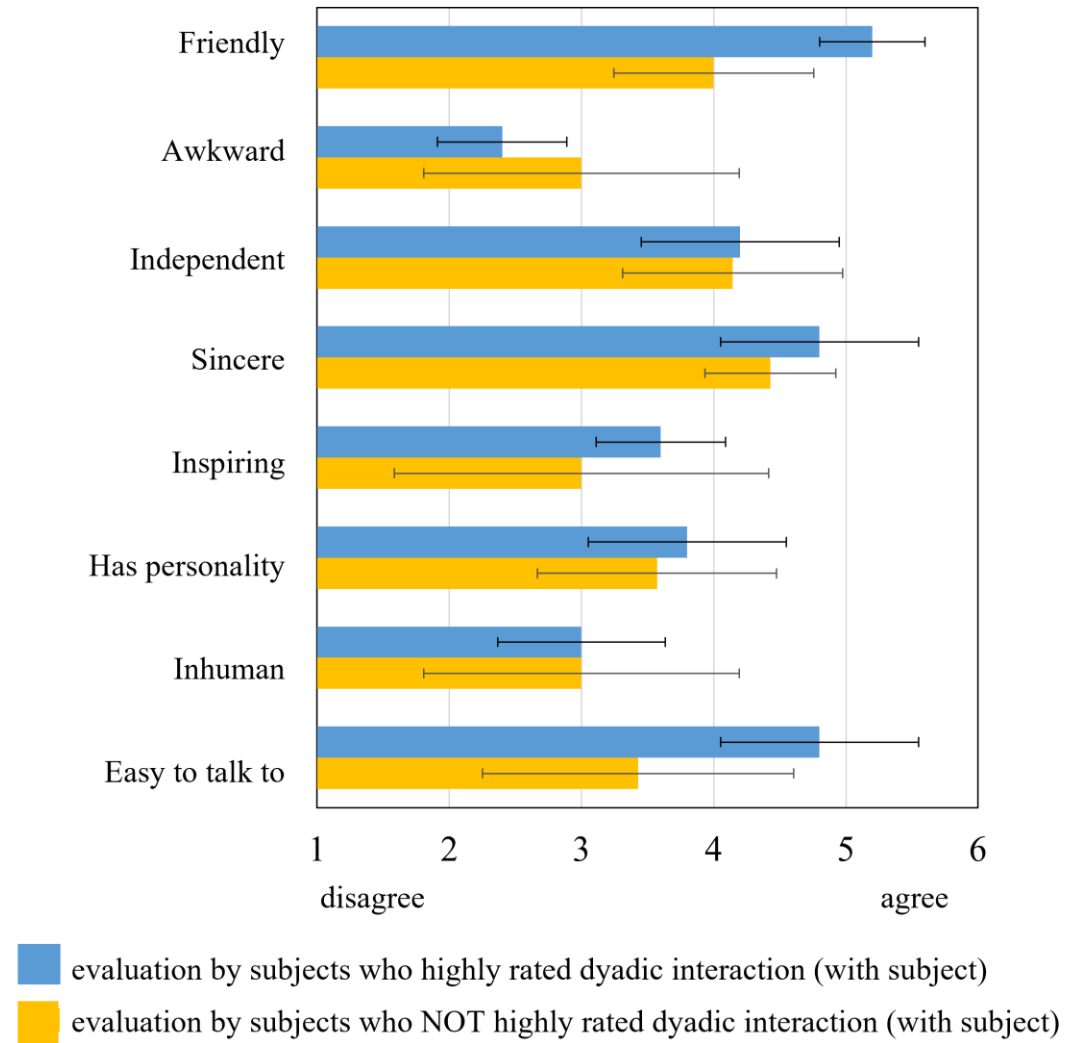


Figure 4.4. The results for the post-dialogue questionnaire regarding the impressions of the dialogue system in the dyadic interaction (with the subject) setting by the dyadic interaction (with the subject) high-rating group and the others.

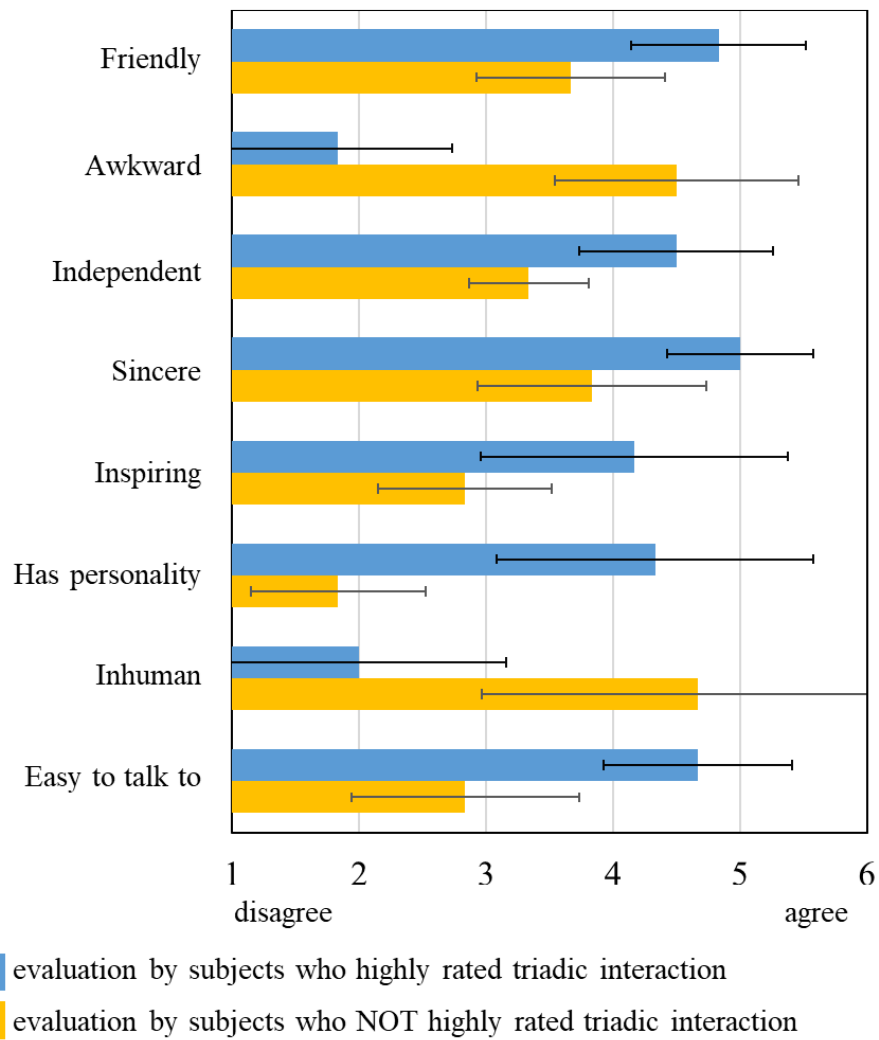


Figure 4.5. The results for the post-dialogue questionnaire regarding the impressions of the dialogue system in the triadic interaction setting by the triadic interaction high-rating group and the others.

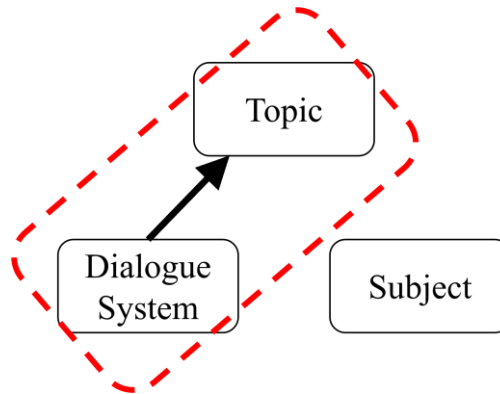
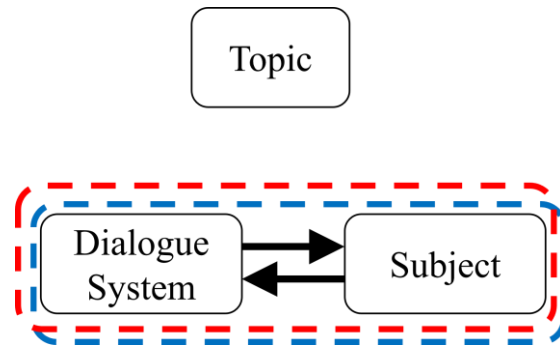


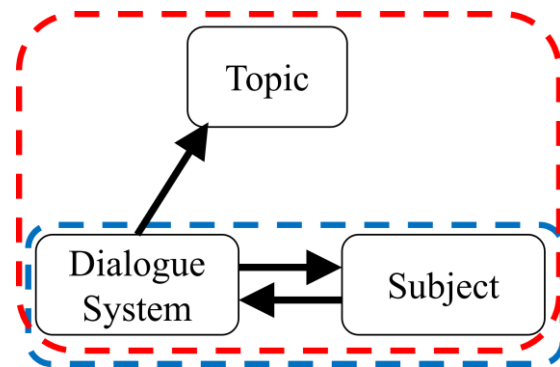
Figure 4.6. The interaction structure that the system constructed without the subject in the dyadic (with the topic) setting.

subjects. On the other hand, the dialogues in the dyadic (with the subject) interaction setting, which is the basic interaction setting, received positive or not negative evaluations. The results are explained based on the successful or failed construction of the interaction structures, depending on the structure that the subject was oriented toward, as follows:

- In the case of the dyadic (with the topic) setting, the structure in the system does not include the subject and does not match any structure to which the subject belongs, as shown in Figure 4.6 with the red dotted line. In this case, the structure construction fails between the system and the subject. In the experiment, the subjects were ignored by the system which talked one-sidedly. In the results, they evaluated the dialogues negatively and commented on the impossibility of dialogue.
- In the case that the subject is oriented toward dyadic interaction, the subject intends to construct the structure as shown in Figure 4.7 with the blue dotted line. This structure matches the structure in the dyadic (with the subject) setting of the system as shown in Figure 4.7(a) with the red dotted line, and they can interact well on the successful construction. However, the structure in the triadic setting of the system extends beyond this structure, as shown in Figure 4.7(b). In this case, the structure construction fails due to the system's undesirable action for the subject. In the experiment, the subjects in the dyadic



(a) In the dyadic interaction (with the subject) setting.



(b) In the triadic interaction setting.

Figure 4.7. The interaction structures desired by the subjects who were oriented toward dyadic interaction (blue dotted line) and the interaction structures in each system setting (red dotted line).

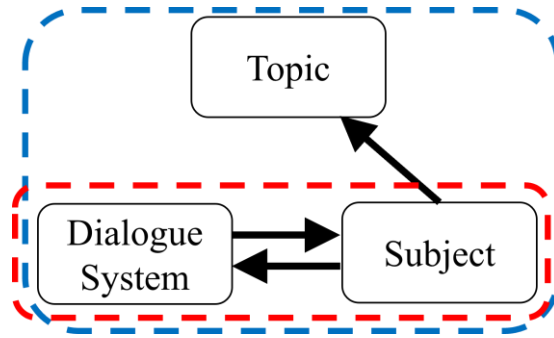
(with the subject) setting high-rating group, would have been oriented toward dyadic interaction. In the results, they evaluated the dialogue in the dyadic (with the subject) setting positively and the dialogue in the triadic setting negatively. Their comments in the triadic setting included discomfort with the dialogue system's inflexibility in persisting with the topic and forcing them to say something that the system wanted. They evaluated the system in the triadic setting as an awkward and inhuman dialogue partner.

- In the case that the subject is oriented toward triadic interaction, the subject intends to construct the structure as shown in Figure 4.8 with the blue dotted line. This structure includes the structure in the dyadic (with the subject) setting of the system as shown in Figure 4.8(a) with the red dotted line and matches the structure in the triadic setting as shown in Figure 4.8(b) with the red dotted line. When the structure is included, the structure construction succeeds even if it is something lacking for the subject, and they can interact relatively well on the successful construction. In the experiment, the subjects in the triadic setting high-rating group, would have been oriented toward triadic interaction. In the results, they evaluated the dialogue in the dyadic (with the subject) setting not negatively and the dialogue in the triadic setting positively. In their comments in the triadic setting, they mentioned that they felt they were listened to attentively.

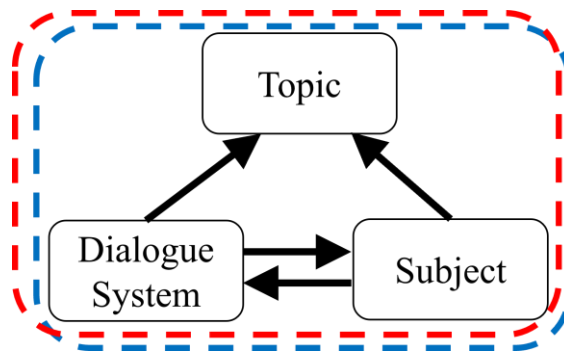
Positive evaluations were based on well-constructed interaction structures. The system's inappropriate behavior in failed structures decreased the quality of the dialogues and damaged the impressions of the system. The interaction structures in the system settings should match the structures intended by the subjects, whether the structures were dyadic or triadic. Under the matching and successful construction, the system fully demonstrated its dialogue capability and behaved pleasantly for the subjects.

The triadic interaction in this experiment involved sharing interest with others. The system in the triadic interaction setting attempted to cooperatively discuss the topic. Thus, whether the subjects intended the triadic or dyadic interactions can be rephrased as whether the subjects wanted to cooperate toward the common aim. When the subjects wanted the system to cooperate in their interaction, they intended triadic interactions. When they did not, their interactions were naturally dyadic. The correct behavior that the system had to choose differed in each case. Avoiding unintended relationships is also important. It is necessary to identify whether cooperation is needed in interaction.

By switching interaction structures to adapt to users' demands, the behavior of systems can become more appropriate for users. For example, if a user wants to be listened to for relaxation,



(a) In the dyadic interaction (with the subject) setting.



(b) In the triadic interaction setting.

Figure 4.8. The interaction structures desired by the subjects who were oriented toward triadic interaction (blue dotted line) and the interaction structures in each system setting (red dotted line).

this user wants to interact in a dyadic interaction structure between just the user and the system, without focusing on any topic seriously. The system should behave in a dyadic setting as a casual listener following casual changes in topics. If a user attempts to discuss a topic to solve a problem, this user wants to interact in a triadic interaction structure between the user, the system, and the topic. The system should behave in a triadic setting for a positive evaluation as an active listener.

In some cases, being fixed in a simple and unbreakable structure can be a better choice for systems. The setting of the basic dyadic interaction structure, which is easy to construct, resulted in moderately positive evaluations from most of the subjects. On the other hand, the setting of the advanced triadic interaction structure was effective in improving the evaluations when the construction succeeded. Superior evaluations can be expected in more advanced structure settings. The structures including additional social factors, for example, the structure in relationships between the user, the other, and the system that supports their interaction, could be useful for further development of systems.

Systems should choose better responses in diverse contexts, that is, socially correct responses adapting to partners' intentions and expectations. As appropriate structures can improve behavior of systems overall, the perspective of choosing the desired interaction structure would help in responding to situations. The social communication capability of dialogue systems indicated in this chapter consists of two stages. First, systems are equipped with applicable interaction structures as behavioral patterns. Second, systems select a suitable structure according to users' choice or identification of users' needs. Systems' appropriate behavior within adapted interaction structures enables the building of better relationships.

The pleasantness of the system's behavior was also reflected in the evaluations of the system. This is discussed in further detail in the next chapter from the perspective of individual preferences.

Chapter 5

Generative AI-based Dialogues

Considering Individual Preferences

This chapter focuses on individual preferences in system behavior to improve personal relationship compatibility between users and systems. The properties of the dialogue system, which were evaluated through the questionnaire items, are investigated to clarify their effects on the dialogue evaluations for each subject.

The communication experiment resulted in diverse subjective evaluations among individual subjects. The links between the evaluations of the system and the dialogue were analyzed for each subject using hierarchical cluster analysis. Based on the analysis results, the properties of the dialogue system were classified into two types: generally and individually relational to positive evaluations of the dialogue. The former included inspiration, a sense of security, and collaboration, while the latter included a sense of distance, personality, and seriousness. Since the latter properties were evaluated based on individual preferences, they should be adjusted to individuals.

5.1 Methodology

As interpersonal preferences vary widely, relationship compatibility is an important factor for dialogue systems with human-like linguistic capability. To build favorable relationships, system properties enhancing relationship compatibility between users and systems should be clarified to adapt to individual preferences.

In the communication experiment, the dialogue system behaved differently depending on each interaction setting, such as listening or ignoring, and compliant or independent. The differently behaving system made different impressions, and the subjects variously evaluated the system

through the questionnaire items. Additionally, even the same behavior of the system in the dialogues was sometimes received either positively or negatively by the subjects. The variations in the evaluations of the system among the subjects reflected individual preferences.

The subjects' evaluations indicated links between the impressions of the system and the qualities of the dialogues. When a dialogue quality was evaluated positively by a subject, some impressions of the system were evaluated highly at the same time. The links are analyzed using hierarchical cluster analysis for each subject, regarding which items for system evaluations were related to positive dialogue evaluations. Based on the analysis results, the evaluation items in the questionnaire are classified. The classification results are discussed for generally and individually favorable design of system properties. Additionally, a part of the evaluation items was developed with reference to research on human friends. These items might be perceived differently by the subjects when applied to the system. Therefore, the interpretations of the evaluation items are also considered.

5.2 Experimental Results

All the subjects interacted with the system and evaluated the system and the dialogues through the post-dialogue questionnaire. The evaluation results are presented from the perspectives of overall trends and individual subjects.

5.2.1 Questionnaire Results

The average and standard deviation of the results for the post-dialogue questionnaire regarding the dialogue system evaluations and the dialogue evaluations are shown in Figure 5.1. The dialogue system in the dyadic interaction (with the topic) setting ignored the subjects in the dialogues, and the evaluations were negative on average. In the dyadic interaction (with the subject) setting and in the triadic interaction setting, dialogues took place between the subjects and the system, and the evaluations were relatively positive on average. On the other hand, each of the subjects gave various evaluations of the system and the dialogues across the three settings. The evaluation results for one subject are shown in Table 5.1 as an example, and the evaluation results for all subjects are shown in Appendices.

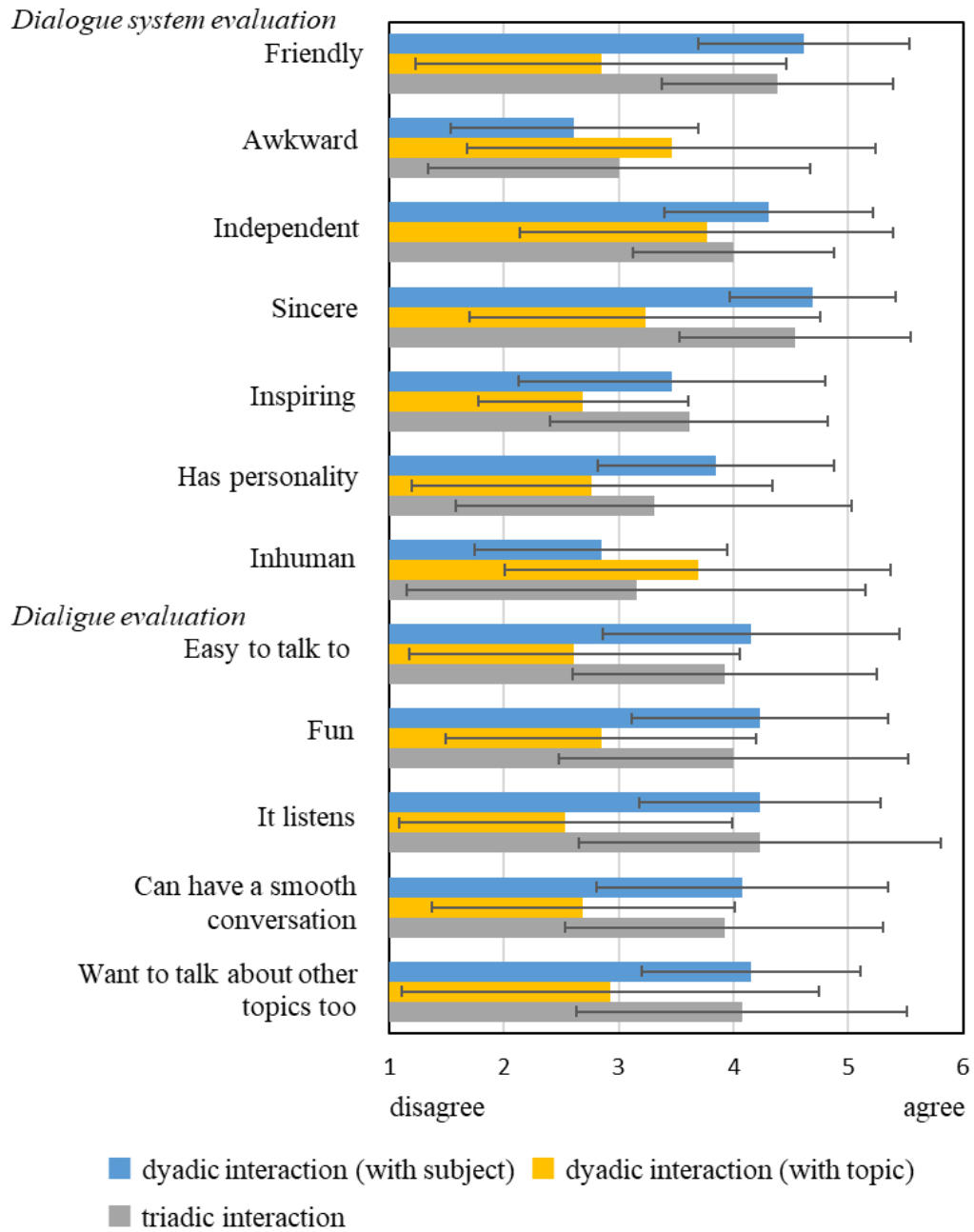


Figure 5.1. The results for the post-dialogue questionnaire regarding the dialogue system evaluations and the dialogue evaluations.

Table 5.1. Questionnaire responses for one of the subjects.

Evaluation Item		Setting		
		Dyadic Interaction (with Subject)	Dyadic Interaction (with Topic)	Triadic Interaction
System	Friendly	5	3	4
	Awkward	3	5	3
	Independent	4	4	4
	Sincere	4	4	5
	Inspiring	5	2	5
	Has personality	5	2	2
	Inhuman	2	5	4
	Easy to talk to	5	2	5
Dialogue	Fun	4	2	5
	It listens	4	2	6
	Can have a smooth conversation	4	2	4
	Want to talk about other topics too	4	2	4

5.3 Analysis and Classification

Hierarchical cluster analysis was conducted on each subject. The analysis results were aggregated to investigate the cumulative results on all subjects. Next, the system evaluation items in the questionnaire were classified according to the cumulative results.

5.3.1 Analysis of Links Between Evaluations of the System and the Dialogue

The post-dialogue questionnaire results were analyzed for each subject to investigate the individual evaluations. Hierarchical cluster analysis with the Ward method was used to examine the links between the evaluations for items about the system and the evaluation for each item about the dialogue. Figure 5.2 shows the dendrograms from the analysis for one subject's evaluations (from Table 5.1) as an example, and the dendrograms for all subjects are shown in Appendices. The strength of the links was quantified with hierarchical closeness on the dendrograms from the analysis. The system evaluation items at the same hierarchy on the dendrograms with a dialogue evaluation item were assigned a step number of 0, items merged at one hierarchical level were assigned a step number of 1, those at two hierarchical levels were assigned a step number of 2, and so forth. A smaller number of steps means a stronger link. The step numbers are shown in brackets after each item in Figure 5.2.

For the subject, as shown in Figure 5.2, the following system evaluations were most strongly linked to each dialogue evaluation. When the system was “Easy to talk to” and “Inspiring”, the dialogues were “Fun”. When the system was “Easy to talk to”, “Inspiring”, and “Friendly”, the dialogues were “It listens”. When the system was “Friendly”, the dialogues were “Can have a smooth conversation” and “Want to talk about other topics too”. The other system evaluations were linked more weakly.

The analysis results for all 13 subjects were aggregated to create the graphs in Figure 5.3, which show the distribution of the subjects for each number of steps. These graphs indicate the number of subjects who linked each system evaluation to each dialogue evaluation and the strength of these links. The color coding of the graphs in Figure 5.3 is based on the distribution shapes, which are explained in the next subsection.

In the remainder of this chapter, a link with a strength of 0 or 1 step is considered to be relational, and a link with a strength of 2 or more steps is non-relational. A step number of 0

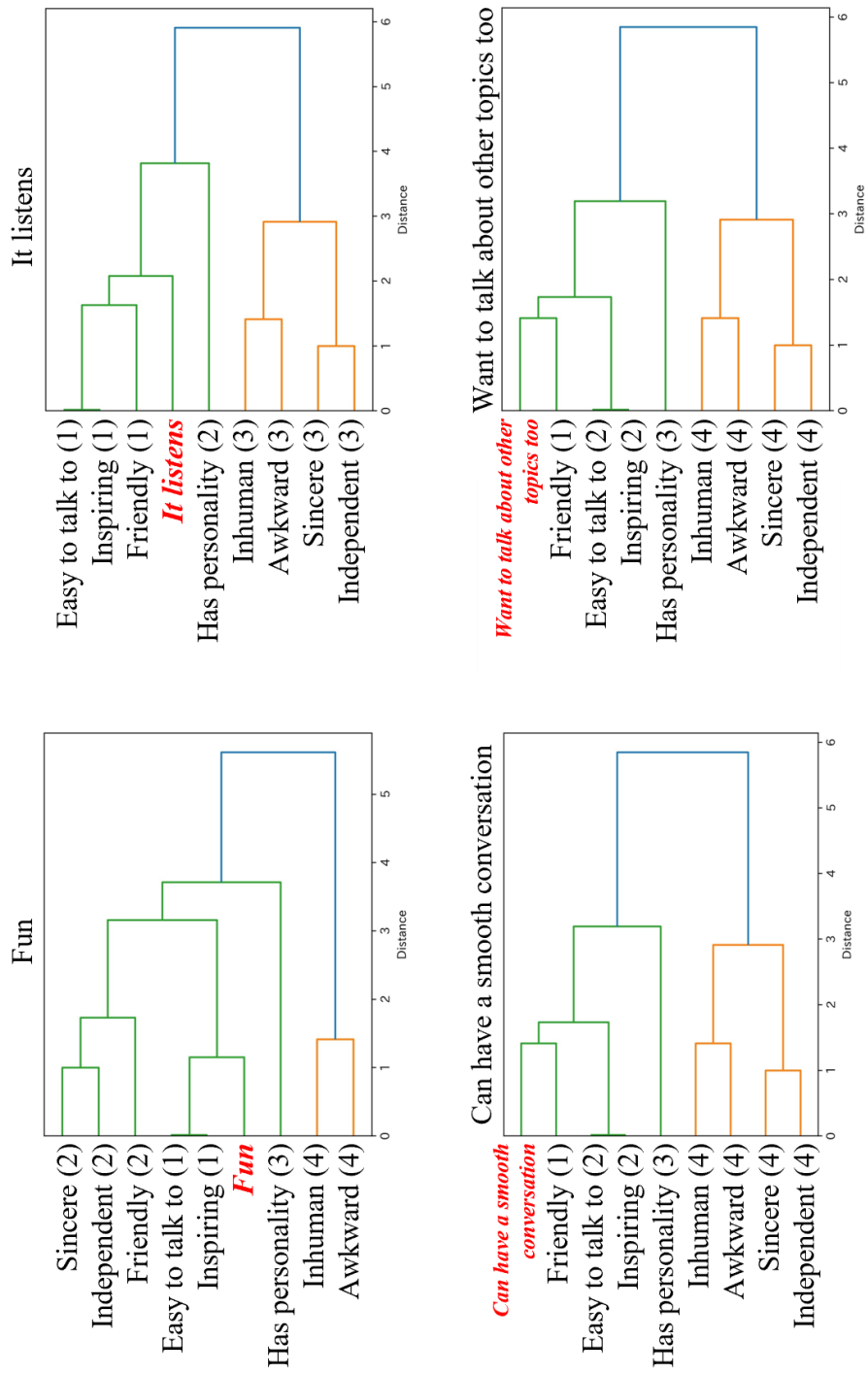


Figure 5.2. Dendrograms from hierarchical cluster analysis for one of the subjects with the step number in brackets after each item.

means that the evaluation values for the two items were completely connected, and a step number of 1 means that the values were not the same but were close. A link with a strength of 2 or more steps is considered weak in comparison to these. According to this distinction, the items for system evaluation that were strongly linked to the dialogue evaluation are treated as relational items, while the others are treated as non-relational items. For example, in the graph of “Friendly” in the “Fun” section, four subjects link “Friendly” for the system to “Fun” for the dialogue with a strength of 0 or 1 step, and the other nine subjects link them with a strength of 2 or more steps. This graph indicates that “Friendly” for the system and “Fun” for the dialogue were relational for four subjects, while they were non-relational for the other nine subjects.

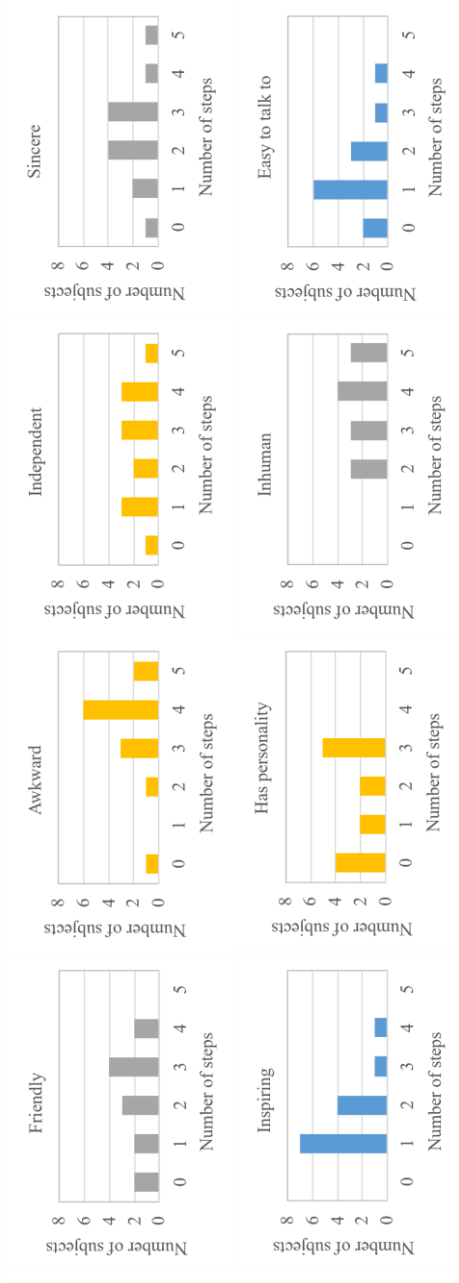
As shown in Figure 5.3, the distributions of subjects in the “Awkward” and “Inhuman” graphs are all at 2 or more steps, except for one subject in each item. While “Awkward” and “Inhuman” functioned as evaluation items for the system as shown in Figure 5.1, the results of the analysis did not show distinct relationships with the dialogue evaluation. The reason would be that these items were negative expressions for a dialogue partner and were difficult to link with the positive dialogue evaluation in the questionnaire. These also showed that the possibility of a positive evaluation of awkward or inhuman behavior was denied. Additionally, since the impressions of words cannot easily be reversed, it is decided not to use reverse scoring values. An exceptional result of “Awkward” is for one subject at 0 steps in the “Fun” section. The subject may have perceived “Awkward” as humorous and linked it to “Fun”. An exceptional result of “Inhuman” is for one subject at 1 step in the “Can have a smooth conversation” section. The subject may have perceived “Inhuman” as neutral and linked it to “Can have a smooth conversation”. Although there are these possible relationships, the exceptional results are too small. The results on “Awkward” and “Inhuman” are excluded from the following classification as these items have no relationship to the dialogue evaluation.

5.3.2 Classification of the System Evaluation Items

The system evaluation items were classified according to the types of relationships with each dialogue evaluation. The classification consists of four categories whether an item is relational or non-relational and whether the relationship is general or individual, based on the characteristics of the subject distributions in Figure 5.3.

The shapes of the distributions in each graph of Figure 5.3 have one or two peaks. The peak positions indicate the evaluation tendencies of the subject groups that are distributed around the

Fun



It listens

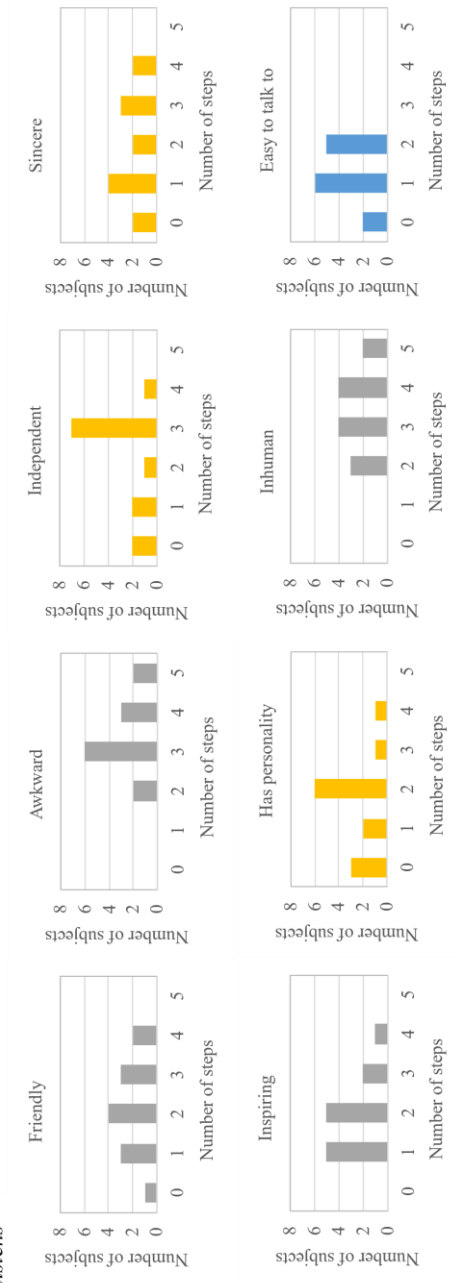


Figure 5.3. The distribution of the subjects for each number of steps from the dendrograms (“Fun” and “It listens”).

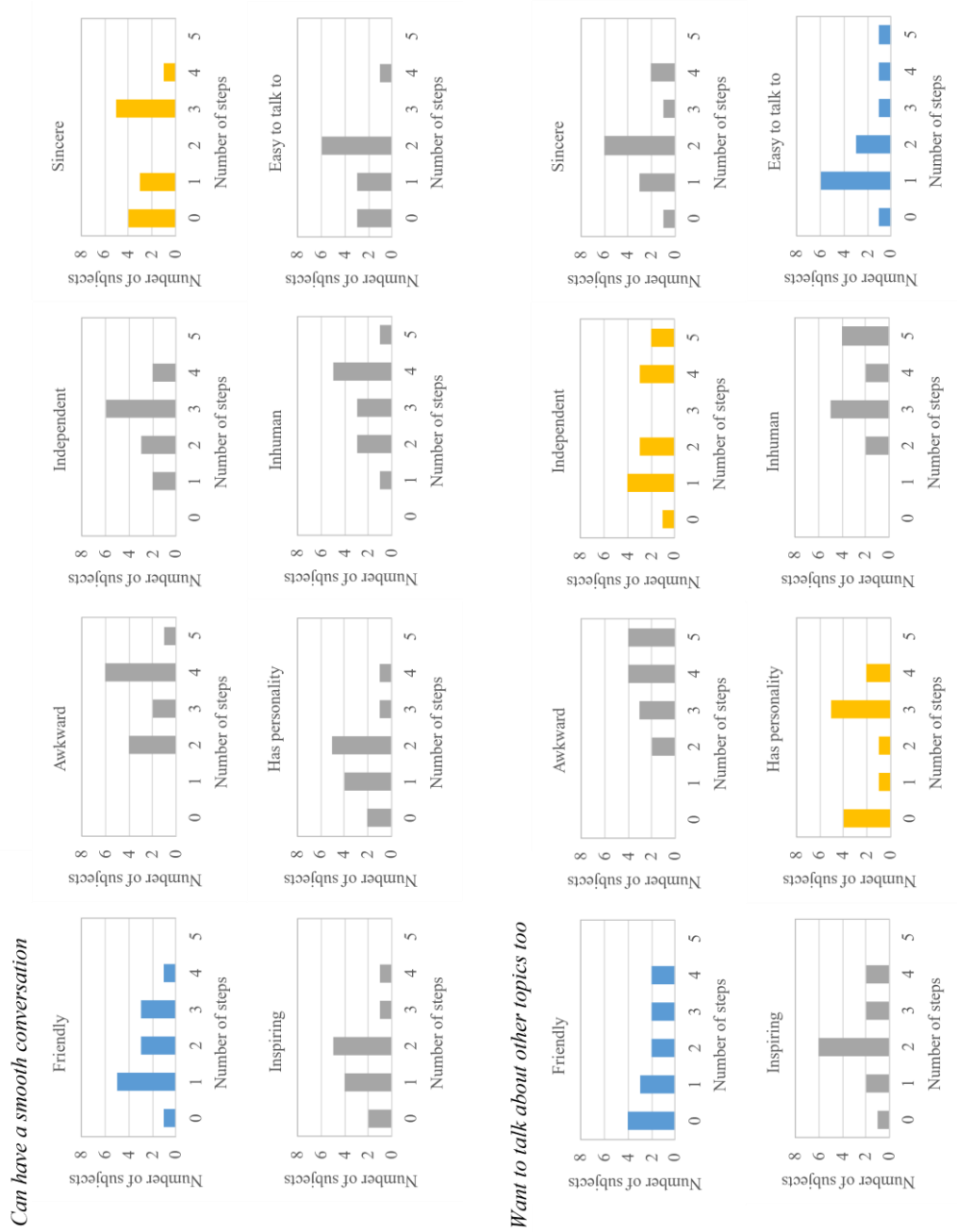
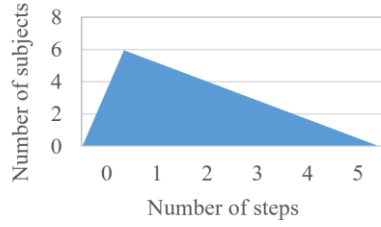


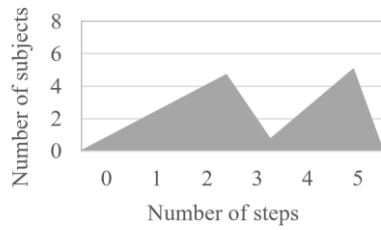
Figure 5.3. The distribution of the subjects for each number of steps from the dendrograms
 (“Can have a smooth conversation” and “Want to talk about other topics too”).



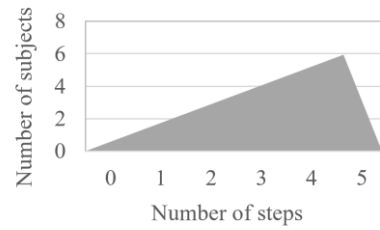
(a) One peak at 0 or 1 step.



(b) Two peaks: one at 0 or 1 step
and the other at 2 or more steps.



(c) Two peaks: both at 2 or more steps.



(d) One peak at 2 or more steps.

Figure 5.4. The distribution patterns.

peaks. When there is one peak, the subjects form one group with the same evaluation tendency. When there are two peaks, the subjects form two groups with different evaluation tendencies. When a peak is at 0 or 1 step, the subject group around this peak tends to relate the system evaluation to the dialogue evaluation. When a peak is at 2 or more steps, the subject group around this peak tends not to relate them. Figure 5.4 shows the distribution patterns, simplified for illustration, based on the number and positions of peaks. If one of the two peaks is at 0 or 1 step and the other peak is at 2 or more steps, as shown in Figure 5.4(b), one subject group tends to relate the system evaluation to the dialogue evaluation, while the other group tends not to relate them.

The color coding in Figure 5.3 corresponds to that used in Figure 5.4. The distribution of

Table 5.2. The classification of the system evaluation items.

Dialogue Evaluation	Generally Relational	Individually Relational	Non-Relational	Non-Relational
Fun	Inspiring Easy to talk to	Independent Has personality		Friendly Sincere
It listens	Easy to talk to	Independent Sincere Has personality		Friendly Inspiring
Can have a smooth conversation	Friendly	Sincere	Easy to talk to	Independent Inspiring Has personality
Want to talk about other topics too	Friendly Easy to talk to	Independent Has personality	Sincere	Inspiring

subjects on “Inspiring” in the “It listens” section in Figure 5.3 has a peak at 1 and 2 steps with equal values. It is determined that there is one peak at 2 steps, matching pattern (d) in Figure 5.4, because the sum of the number of subjects at 2 or more steps is larger than that at 0 or 1 step.

The classification results of the six evaluation items, except for “Awkward” and “Inhuman”, are shown in Table 5.2. Items with one peak at 0 or 1 step, as shown in Figure 5.4(a), were classified into the first column. Items with two peaks, one at 0 or 1 step and the other at 2 or more steps, as shown in Figure 5.4(b), were classified into the second column. Items with two peaks both at 2 or more steps, as shown in Figure 5.4(c), were classified into the third column. Items with one peak at 2 or more steps, as shown in Figure 5.4(d), were classified into the fourth column.

The items in the first column of Table 5.2 are generally relational items. The subjects formed one group and tended to relate the system evaluations for the items to the dialogue evaluation. The items in the second column are individually relational items. The subjects formed two groups, and the evaluation tendencies differed across subjects for the items. Some subjects in one of the two groups related the evaluations, and the others did not. The items in the third and fourth columns are non-relational items. In this study, the items in the third column are treated as simply

non-relational, along with the items in the fourth column. Although there were two groups of subjects and they may have differed in their evaluations for the items in the third column, neither group related the evaluations. Since the aim is to investigate how the subjects related the evaluations, how they did not relate them is not addressed. Further research on negative evaluations, such as indifference or negative impacts, could provide additional findings.

5.4 Discussion

In this experiment, the subjects communicated with the human-like dialogue system using generative AI. The system behaved differently in three interaction settings. The subjective evaluations of the system and the dialogues varied among individual subjects. Focusing on the individual evaluations, a cluster analysis of the links in the evaluations was conducted. Based on the analysis results, the cumulative results of all subjects' evaluations were organized, as shown in Figure 5.3. According to the cumulative results, the system evaluation items were classified, as shown in Table 5.2. The “awkward” and “inhuman” items were excluded from the classification since they were not related to the positive dialogue evaluations. The irrelevance of “inhuman” indicates that the human-likeness of the system was preferred in the dialogues.

To identify the system properties that influenced the subjects' evaluations, the items in Table 5.2 are discussed. The subjects' interpretations of the item words are considered with the premise that the system is not human.

The system evaluation items in Table 5.2:

- The “Inspiring” item was developed to evaluate an attractiveness factor of a friend that provides positive inspiration [72]. The “Inspiring” item was only related to the “Fun” from the dialogue category. In other words, dialogues with an inspiring partner were fun. This is a reasonable result of a literal interpretation of “Inspiring”. The property of inspiration was evaluated through the item “Inspiring”.
- The “Easy to talk to” item was developed to evaluate a sense of security. An interpersonal relationship with a sense of security means a relaxed relationship without worries or barriers. As shown in previous research on friendships, the sense of security is the most fundamental attractiveness factor of a friend, and the survey item “Easy to talk to” best reflects this [72]. The “Easy to talk to” item was related to three dialogue evaluations, except for “Can have a

smooth conversation”. The result indicates that the item was most commonly related to positive evaluations, as it was the most frequent and generally relational among the subjects. It is also natural that a sense of security is a fundamental element of the relationship with the system. Similarly to interpersonal relationships, the property of a sense of security was evaluated through the item “Easy to talk to”.

- The “Friendly” item was developed with the expectation that a friendly system would be evaluated positively as a dialogue partner. The “Friendly” item was related to “Can have a smooth conversation”, which indicates linguistic abilities, and “Want to talk about other topics too”, which indicates a wish to continue the relationship, but not to “Fun” or “It listens”, which involve feelings. It seems that the “Friendly” item was interpreted as being collaborative, which refers to the system’s dialogue capability to make dialogue easy for the subjects. Compared with being friendly, being collaborative is a more plausible expectation for the system. The property of collaboration was evaluated through the item “Friendly”.
- The “Independent” item was developed to evaluate a sense of distance, an attractiveness factor of a friend that maintains a moderate distance [72]. The “Independent” item was related to “Fun”, “It listens”, and “Want to talk about other topics too” among some subjects. In other words, dialogues with an independent partner made some subjects feel good and caused a wish to continue. This is a reasonable result for an interpretation of “Independent” as maintaining a moderate distance, similar to the case of human friends. The property of a sense of distance was evaluated through the item “Independent”.
- The “Has personality” item was developed with reference to a survey about impressions of robots [73]. The words “Has personality” are reasonably interpreted literally regarding systems. The “Has personality” item was related to three dialogue evaluations, except for “Can have a smooth conversation”, among some subjects. It is natural that some subjects preferred a partner with a personality. The property of personality was evaluated through the item “Has personality”.
- The “Sincere” item was developed to evaluate the attractiveness factor of a friend. Sincerity is an important factor in lasting friendships [72]. However, the “Sincere” item was not related to “Want to talk about other topics too”, which indicates a wish to continue the relationship. It was related to “Can have a smooth conversation” and “It listens” among some subjects, but not to “Fun”. When the dialogue system listened well and responded fluently, the system was evaluated as sincere. Unlike in interpersonal relationships, “Sincere” in the system was interpreted in terms of reliability rather than trust. Considering that an attentive attitude and

conversational ability were evaluated individually through the word sincere, the “Sincere” item is referred to as seriousness. The property of seriousness was evaluated through the item “Sincere”.

The properties of the system related to the positive dialogue evaluations included inspiration, a sense of security, collaboration, a sense of distance, personality, and seriousness.

These properties are divided into two types, generally and individually relational, corresponding to the first and second columns of Table 5.2. The three properties of inspiration, a sense of security, and collaboration generally received positive evaluations. The other three properties of a sense of distance, personality, and seriousness had different effects on individuals. Equipping with the former properties is expected to make dialogues more comfortable for people. Adjusting the latter properties is expected to make dialogues more compatible for individuals.

In particular, to adopt individual preferences, personalizing properties of individually relational type would be effective. For example, tuning the level of interference with users’ actions, characterization, and the frequency of repeated confirmations. On the other hand, for general use, modest designs in these individually relational properties are predicted to reduce the risk of the system being disliked. This prediction is consistent with the results of previous research as follows. Although there are individual differences, the setting supported by a larger number of people involves avoiding small talk and maintaining a formal tone, identifying itself as a bot, and asking follow-up questions [40], that is, maintaining a formal distance, having less personality, and expressing a serious attitude. A dialogue system in a highly serious setting, with a single focused topic, receives both positive and negative evaluations, while a system in a less serious setting, with no focused topic, receives positive evaluations generally, as indicated in Chapter 4. When diagnosing a disease through medical interviews, patients are more likely to self-disclose honestly with a virtual anonymous agent, which has less personality, than with a human specialist about some topics [63].

The properties of the dialogue system were classified and examined. Although favorable properties of dialogue systems have been approached mainly by analogy with personality traits in humans, it was clarified that they differ from humans in several respects. Dialogue systems can be developed to have their own unique advantages. It was also revealed that some properties have no generally correct setting beyond individual adjustments. The approach in this chapter to address system properties considering individual preferences would help to identify what properties should be targeted for adjustment of dialogue systems.

Chapter 6

Conclusions

6.1 Conclusions of This Thesis

As the linguistic capability of dialogue systems improves, their roles are going beyond merely transmitting information. Advanced dialogue systems are expected to support people personally through linguistic communication. The importance of how systems interact with humans and build supportive relationships is increasing, and social adaptability and personal compatibility issues similar to those in human relationships emerge. This thesis investigated subjective evaluations of interactions with the generative AI-based dialogue system to improve relationships between humans and dialogue systems.

The dialogue system was developed to communicate with the subjects in natural language via voice, featuring a human-like facial expression function. These communication abilities were achieved with generative AI. The system behaved differently in interactions with the subjects depending on three interaction settings. The one-to-one dyadic interaction, triadic interaction with joint attention to a topic, and non-communicative settings were designed with reference to the developmental stages of children's social communication ability. The dialogues were both successful and failed, and the subjective evaluations of the system and the dialogues were conducted through the questionnaire. The evaluations resulted in various responses.

The dialogue evaluations were discussed based on the interaction structures, that is, the positions the system took in the interactions. As a result, positive evaluations were based on well-constructed structures. The system's inappropriate behavior in failed structures decreased the quality of the dialogues and worsened the evaluation of the system. The interaction structures in the system settings needed to match the structures intended by the subjects, whether the structures were dyadic or triadic. Under the matching and successful construction, the system fully demonstrated its dialogue capability and behaved pleasantly for the subjects. By switching

interaction structures in systems to adapt to users' demands, the structures can be successfully constructed between users and systems, and the behavior would become more appropriate for users. Socially correct behavior, adapting to social contexts on partners' intentions and expectations, affects users' evaluations similarly to interpersonal relationships. Dialogue systems could build better relationships through appropriate interactions.

The evaluations of the system and the dialogues were analyzed focusing on individual preferences. The analysis results showed that human-like behaviors were preferred in the dialogues and the system properties were classified into two types. Properties of inspiration, a sense of security, and collaboration generally improved the dialogue evaluations. Properties of a sense of distance, personality, and seriousness individually affected the dialogue evaluations depending on each subject. Properties of individually relational type need to adapt to users' individual preferences. A design approach to system properties in accordance with individuality would be useful for making human-like dialogue systems more compatible with users.

Dialogue systems with human-like linguistic capability are required to participate in users' interpersonal relationships in order to support human life through linguistic communication. In this thesis, how systems should behave to build relationships with humans was explored as the start of their relationships. It was clarified that adjusting interaction structures is effective for adapting to social contexts, and adjusting system properties is effective for improving relationship compatibility. Appropriate and favorable behaviors of systems can be achieved with these adjustments. These findings contribute to building better relationships with advanced human-like dialogue systems and realizing human well-being supported by future systems.

6.2 Limitations and Notes

The sample size of 13 subjects in the experiment was small for statistical verification, and the reliability of the results is limited. The effects of interaction structures in the generative AI-based system were demonstrated to some extent, and the findings regarding system properties are able to explain the participants' responses from previous research. However, further verification is needed.

Furthermore, it is necessary to note the risks associated with AI dialogue systems. AI technology has been discussed in terms of social and ethical risks that may undermine people's

autonomy, privacy, and equity. Systems that include AI technology need to be developed ethically and safely for users [74, 75, 76, 77]. Human-like AI dialogue systems, adapted to persons, will become comfortable partners in daily life. However, they also involve these risks, so their development should be ethical.

6.3 Future Work

In future work, some trials are needed to complement this research. To further verify the effectiveness of the findings, a test of a renewed system that adapts to users' desired behaviors in contexts and preferences is planned. Interaction structures including additional participants and social factors are also worth a try. In addition, emotional functions will be examined as an important factor in systems. The emotional expression function of the system was active with mainly positive expressions during the dialogues. Although the effect in non-emotional dialogues was merely that the system was perceived as natural in this research, handling emotions is important in terms of addressing individuals. There would naturally be socially appropriate behavior and individual preferences in the emotional responses of systems. Adapting to users' emotions could improve their relationships. Emotional functions should be explored with affective computing technologies, including human emotion recognition and sentiment analysis, which aim to identify and express emotions and respond intelligently to human emotions [78]. These examinations plan to include quantitative data from temporal and biometric measurements in order to enhance validity. On the other hand, interactive systems targeting specific users, such as children and the elderly, are needed and should be researched [79, 80, 81]. Results different from those in this study could emerge for specific groups, and research concerning these aspects is also required.

Generative AIs continue to evolve. ChatGPT has been updated from 3.5, which was used in the experiment, to 4 and now responds more skillfully. More advanced linguistic capability will refine the responses of systems and provide participants with enjoyable conversations. However, OpenAI announced that they reverted the update to GPT-4o due to issues with overly agreeable responses on April 29, 2025. Its sycophantic behavior was disliked by users. In social interactions, behavior merely following others is a basic and low-level attitude. The social abilities to use appropriate words and actions in relationships are required. Damage to relationships between the

subjects and the dialogue system due to the lack of such abilities was observed in this research as well. Not only linguistic abilities but also social abilities are essential for future systems. Further research is necessary to explore the favorable behavior and the design with social abilities for supportive relationships between humans and systems.

Bibliography

- [1] Følstad, A., Skjuve, M., and Brandtzaeg, P. B.: Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design. *Internet Science*, 2019, 11551, 145–156. https://doi.org/10.1007/978-3-030-17705-8_13.
- [2] Reeves, B. and Nass, C.: *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, 1996.
- [3] Ho, A., Hancock, J., and Miner, A. S.: Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations with a Chatbot. *Journal of Communication*, 2018, 68, 712–733. <https://doi.org/10.1093/joc/jqy026>.
- [4] Okada, Y., Kimoto, M., Iio, T., Shimohara, K., and Shiomi, M.: Two is Better than One: Apologies from Two Robots are Preferred. *PLoS ONE*, 2023, 18, e0281604. <https://doi.org/10.1371/journal.pone.0281604>.
- [5] 福田健: 聞き上手になれる本. 日本実業出版社, 1983.
- [6] 一般財団法人 NHK 放送研修センター・日本語センター: 誰からも好かれる NHK の話し方. KADOKAWA, 2016.
- [7] 山田ズーニー: あなたの話はなぜ「通じない」のか. 筑摩書房, 2003.
- [8] 東畑開人: 聞く技術 聞いてもらう技術. 筑摩書房, 2022.
- [9] STEVEN A. COLE, JULIAN BIRD: メディカルインタビュー 第2版: 三つの機能モデルによるアプローチ. メディカル・サイエンス・インターナショナル, 2003.
- [10] 石井均: 糖尿病ケアの知恵袋—よき「治療同盟」をめざして. 医学書院, 2004.
- [11] 河合隼雄: 人の心はどこまでわかるか. 講談社, 2000.
- [12] 山上敏子: 方法としての行動療法. 金剛出版, 2007.
- [13] マイロン F. ワイナー: 人間としての心理治療者: 自己開示の適用と禁忌. 有斐閣, 1983.
- [14] 藤山直樹, 笠井清登, 松木邦裕, 中村紀子, 中原睦美, 伊藤絵美, 村井俊哉, 熊谷晋一郎, 熊倉陽介: ころを使うということ—いま求められる心理職のアイデンティティ. 岩崎学術出版社, 2020.
- [15] 関谷大輝: あなたの仕事、感情労働ですよ？. 花伝社, 2016.
- [16] Tomasello, M.: *Becoming Human: A Theory of Ontogeny*. Belknap Press of Harvard

University Press, 2021.

- [17] 松井豊, 白石崇, 村井潤一郎:対人心理学の視点. ブレーン出版, 2002.
- [18] 長田攻一:対人コミュニケーションの社会学. 学文社, 2008.
- [19] 榎本博明:自己開示の心理学的研究. 北大路書房, 1997
- [20] 三牧陽子:ポライトネスの談話分析—初対面コミュニケーションの姿としくみ. くろしお出版, 2013.
- [21] 大坊郁夫:幸福を目指す対人社会心理学—対人コミュニケーションと対人関係の科学. ナカニシヤ出版, 2012.
- [22] 諏訪きぬ, 戸田有一, 高橋真由美, 上月智晴, 中坪史典:保育における感情労働—保育者の専門性を考える視点として. 北大路書房, 2011.
- [23] 三宮真智子:思考・感情を表現する力を育てるコミュニケーション教育の提案:メタ認知の観点から. 鳴門教育大学学校教育実践センター紀要, 2005, 19, 151–161. <https://doi.org/10.24727/00025014>.
- [24] 齋藤孝:話し上手 聞き上手. 筑摩書房, 2007.
- [25] 相川充:人づきあいの技術: ソーシャルスキルの心理学. サイエンス社, 2009.
- [26] Kosinski, M.:Evaluating Large Language Models in Theory of Mind Tasks. Proceedings of the National Academy of Sciences, 2024, 121, e2405460121. <https://doi.org/10.1073/pnas.2405460121>.
- [27] Jones, C. R. and Bergen, B. K.:People Cannot Distinguish GPT-4 from a Human in a Turing Test. arXiv:2405.08007, 2024. <https://doi.org/10.48550/arXiv.2405.08007>.
- [28] フレデリック・カプラン:ロボットは友だちになれるか. NTT 出版, 2011.
- [29] 岡田美智男:ロボット: 共生に向けたインタラクション. 東京大学出版会, 2022.
- [30] 三野星弥, 吉川雄一郎, 伴碧, 石黒浩:友人グループ内での長期間利用による他者情報のやり取りを行う日常対話チャットボットの評価. 人工知能学会論文誌, 2022, 37, 3, IDS-I_1-14. https://doi.org/10.1527/tjsai.37-3_IDS-I.
- [31] ドナルド A.ノーマン:エモーショナル・デザイン—微笑を誘うモノたちのために. 新曜社, 2004.
- [32] 横山章光:臨床精神医学から見た, ロボット・セラピーの未来～アニマル・セラピーの知見も鑑みて～. 計測と制御, 2012, 51, 7, 598–602. <https://doi.org/10.11499/sicejl.51.598>.
- [33] 吉田たかよし:話し上手は「相づち」が9割. 宝島社, 2007.
- [34] 藤堂健世, 佐久間洋司, 大澤博隆, 清田陽司:AI エージェントの社会実装における論点の整理—「AI さくらさん」の事例から—. 人工知能, 2020, 35, 5, 638–642. https://doi.org/10.11517/jjsai.35.5_638.

- [35] Zhou, H., Huang, M., Zhang, T., Zhu, X., and Liu, B.: Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, 2018, 730–738. <https://dl.acm.org/doi/10.5555/3504035.3504125>.
- [36] Nißen, M., Rüegger, D., Stieger, M., Flückiger, C., Allemand, M., Wangenheim, F. v., and Kowatsch, T.: The Effects of Health Care Chatbot Personas with Different Social Roles on the Client-Chatbot Bond and Usage Intentions: Development of a Design Codebook and Web-based Study. *Journal of Medical Internet Research*, 2022, 24, 4. <https://doi.org/10.2196/32630>.
- [37] Miyanishi, T., Hirayama, J., Kanemura, A., and Kawanabe, M.: Answering Mixed Type Questions About Daily Living Episodes. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, 2018, 4265–4271. <https://doi.org/10.24963/ijcai.2018/593>.
- [38] 東中竜一郎: AI の雑談力. KADOKAWA, 2021.
- [39] Rapp, A., Curti, L., and Boldi, A.: The Human Side of Human-Chatbot Interaction: A Systematic Literature Review of Ten Years of Research on Text-based Chatbots. *International Journal of Human-Computer Studies*, 2021, 151, 102630. <https://doi.org/10.1016/j.ijhcs.2021.102630>.
- [40] Kätsyri, J., Förger, K., Mäkräinen, M., and Takala, T.: A Review of Empirical Evidence on Different Uncanny Valley Hypotheses: Support for Perceptual Mismatch as One Road to the Valley of Eeriness. *Frontiers in Psychology*, 2015, 6. <https://doi.org/10.3389/fpsyg.2015.00390>.
- [41] Svenningsson, N. and Faraon, M.: Artificial Intelligence in Conversational Agents: A Study of Factors Related to Perceived Humanness in Chatbots. Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference, 2019, 151–161. <https://doi.org/10.1145/3375959.3375973>.
- [42] 貴志悠, 神田智子: 対話エージェントのうなずきタイミングが発話長に及ぼす影響分析. HAI シンポジウム 2011, 2011, I-1B-4.
- [43] 片山晋, 米澤拓郎, 河口信夫: テキスト対話システムにおける感情制御を伴う応答生成手法. *知能と情報*, 2022, 34(3), 568–578. https://doi.org/10.3156/jsoft.34.3_568.
- [44] Durante, Z., Huang, Q., Wake, N., Gong, R., Park, J. S., Sarkar, B., Taori, R., Noda, Y.,

- Terzopoulos, D., Choi, Y., Ikeuchi, K., Vo, H., Fei-Fei, L., and Gao, J.: Agent AI: Surveying the Horizons of Multimodal Interaction. arXiv:2401.03568, 2024. <https://doi.org/10.48550/arXiv.2401.03568>.
- [45] Zhang, J., Oh, Y. J., Lange, P., Yu, Z., and Fukuoka, Y.: Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint. *Journal of Medical Internet Research*, 2020, 22, 9. <https://doi.org/10.2196/22845>.
- [46] Ayedoun, E., Hayashi, Y., and Seta, K.: Adding Communicative and Affective Strategies to an Embodied Conversational Agent to Enhance Second Language Learners' Willingness to Communicate. *International Journal of Artificial Intelligence in Education*, 2019, 29, 29–57. <https://doi.org/10.1007/s40593-018-0171-6>.
- [47] Herrmann-Werner, A., Loda, T., Junne, F., Zipfel, S., and Mamlouk, A. M.: “Hello, my name is Melinda”—Students' Views on a Digital Assistant for Navigation in Digital Learning Environments; A Qualitative Interview Study. *Frontiers in Education*, 2021, 5. <https://doi.org/10.3389/educ.2020.541839>.
- [48] 植田一博: アニマシー知覚: 人工物から感じられる生物らしさ. *日本ロボット学会誌*, 2013, 31, 9, 833–835. <https://doi.org/10.7210/jrsj.31.833>.
- [49] Gao, T., McCarthy, G., and Scholl, B. J.: The Wolfpack Effect: Perception of Animacy Irresistibly Influences Interactive Behavior. *Psychological Science*, 2010, 21, 1845–1853. <https://doi.org/10.1177/0956797610388814>.
- [50] Kanakogi, Y., Okumura, Y., Inoue, Y., Kitazaki, M., and Itakura, S.: Rudimentary Sympathy in Preverbal Infants: Preference for Others in Distress. *PLoS ONE*, 2013, 8, 6, e65292. <https://doi.org/10.1371/journal.pone.0065292>.
- [51] 福田玄明, 植田一博: 実際の生物を用いたアニマシー知覚の脳内基盤の検討. *認知科学*, 2011, 18, 1, 64–78. <https://doi.org/10.11225/jcss.18.64>.
- [52] Schultz, R. T., Gauthier, I., Klin, A., Fulbright, R. K., Anderson, A. W., Volkmar, F., Skudlarski, P., Lacadie, C., Cohen, D. J., and Gore, J. C.: Abnormal Ventral Temporal Cortical Activity During Face Discrimination Among Individuals with Autism and Asperger Syndrome. *Archives of General Psychiatry*, 2000, 57, 4, 331–340. <https://doi.org/10.1001/archpsyc.57.4.331>.
- [53] 森口佑介: わたしを律するわたし—子どもの抑制機能の発達. 京都大学学術出版会, 2012.
- [54] 阿部香織, 牧川方昭: 視覚的関心領域の頭部運動への現れの計測. 生体医工学シンポ

- ジウム 2008 講演予稿集, 2008, 166–174.
- [55] アンジェロ・カンジェロシ, マシュー・シュレシンジャー: 発達ロボティクスハンドブック. 福村出版, 2019.
- [56] 小嶋秀樹, 高田明: 社会的相互行為への発達のアプローチ: 社会のなかで発達するロボットの可能性 (特集 社会的相互行為). 人工知能, 2001, 16, 6, 812–818. https://doi.org/10.11517/jjsai.16.6_812.
- [57] Scassellati, B.: Theory of Mind for a Humanoid Robot. *Autonomous Robots*, 2002, 12, 13–24. <https://doi.org/10.1023/A:1013298507114>.
- [58] Ahmad, R., Siemon, D., Gnewuch, U., and Robra-Bissantz, S.: Designing Personality-Adaptive Conversational Agents for Mental Health Care. *Information Systems Frontiers*, 2022, 24, 923–943. <https://doi.org/10.1007/s10796-022-10254-9>.
- [59] 滝僚平, 前田陽一郎, 高橋泰岳: 人とロボットのインタラクティブ情動コミュニケーションにおけるロボットの有効反応モデル. 第 26 回ファジィシステムシンポジウム講演論文集, 2010, MD3-2. <https://doi.org/10.14864/fss.26.0.48.0>.
- [60] 宮本友樹, 永井望, 満田雄斗, 磐下大樹, 遠藤水紀, 鈴木章弘, 片上大輔: Risky Politeness Strategy の提案と雑談対話システムへの実装. 人工知能学会論文誌, 2022, 37, 3, IDS-G_1-16. https://doi.org/10.1527/tjsai.37-3_IDS-G.
- [61] 水上雅博, 杉山弘晃, 有本庸浩, 東中竜一郎, 光田航, 小林哲生: なりきり AI プラットフォームを活用した自治体連携による案内業務対話システム構築. 人工知能学会論文誌, 2023, 38, 3, B-MA2_1-14. https://doi.org/10.1527/tjsai.38-3_B-MA2.
- [62] 小林一樹, 船越孝太郎, 小松孝徳, 山田誠二, 中野幹生: ASE に基づく相槌によるロボットとの対話体験の向上. 人工知能学会論文誌, 2015, 30, 4, 604–612. <https://doi.org/10.1527/tjsai.30.604>.
- [63] Yokotani, K., Takagi, G., and Wakashima, K.: Advantages of Virtual Agents over Clinical Psychologists During Comprehensive Mental Health Interviews Using a Mixed Methods Design. *Computers in Human Behavior*, 2018, 85, 135–145. <https://doi.org/10.1016/j.chb.2018.03.045>.
- [64] Hill, J., Ford, W. R., and Farreras, I. G.: Real Conversations with Artificial Intelligence: A Comparison Between Human-Human Online Conversations and Human-Chatbot Conversations. *Computers in Human Behavior*, 2015, 49, 245–250. <https://doi.org/10.1016/j.chb.2015.02.026>.
- [65] Mou, Y. and Xu, K.: The Media Inequality: Comparing the Initial Human-Human and Human-AI Social Interactions. *Computers in Human Behavior*, 2017, 72, 432–440.

- <https://doi.org/10.1016/j.chb.2017.02.067>.
- [66] Brandtzaeg, P. B. and Følstad, A.: Why People Use Chatbots. *Internet Science*, 2017, 10673, 377–392. https://doi.org/10.1007/978-3-319-70284-1_30.
- [67] Löffler, D., Schmidt, N., and Tscharn, R.: Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018, 334–343. <https://doi.org/10.1145/3171221.3171261>.
- [68] 河原美彩子, 澤田佳子, 田中章浩: ヒューマノイドロボットに対する人間の多感覚的な感情認知. *日本感性工学会論文誌*, 2021, 20, 3, 329–335. <https://doi.org/10.5057/jjske.TJSKE-D-21-00015>.
- [69] 山内厚志, 寺田和憲, 伊藤昭: 動的な発色によるロボットの感情表出. *ヒューマンインタフェース学会論文誌*, 2011, 13, 1, 41–52. https://doi.org/10.11184/his.13.1_41.
- [70] Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., and Schyns, P. G.: Facial Expressions of Emotion are not Culturally Universal. *Proceedings of the National Academy of Sciences*, 2012, 109, 19, 7241–7244. <https://doi.org/10.1073/pnas.1200155109>.
- [71] 浜田寿美男: 「私」とは何か. 講談社, 1999.
- [72] 西浦真喜子, 大坊郁夫: 同性友人に感じる魅力が関係継続動機に及ぼす影響: 個人にとっての重要性の観点から. *対人社会心理学研究*, 2010, 10, 115–123. <https://doi.org/10.18910/10703>.
- [73] Chidori, H. and Matsuzaki, G.: Impression for the Robot that the Purpose of Communication. *Proceedings of the 59th Annual Conference of Japanese Society for the Science of Design*, 2012. https://doi.org/10.11247/jssd.59.0_258.
- [74] Ortega-Bolaños, R., Bernal-Salcedo, J., Ortiz, M. G., Sarmiento, J. G., Ruz, G. A., and Tabares-Soto, R.: Applying the Ethics of AI: A Systematic Review of Tools for Developing and Assessing AI-based Systems. *Artificial Intelligence Review*, 2024, 57, 110. <https://doi.org/10.1007/s10462-024-10740-3>.
- [75] Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., and Coiera, E.: Conversational Agents in Healthcare: A Systematic Review. *Journal of the American Medical Informatics Association*, 2018, 25, 1248–1258. <https://doi.org/10.1093/jamia/ocy072>.
- [76] Sundar, S. S.: Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAI). *Journal of Computer-Mediated Communication*, 2020, 25, 74–88. <https://doi.org/10.1093/jcmc/zmz026>.

- [77] 松尾豊, 西田豊明, 堀浩一, 武田英明, 長谷敏司, 塩野誠, 服部宏充, 江間有沙, 長倉克枝: 人工知能と倫理. 人工知能, 2016, 31, 5, 635–641. https://doi.org/10.11517/jjsai.31.5_635.
- [78] Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., Gao, S., Sun, Y., Ge, W., Zhang, W., and Zhang, W.: A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances. Information Fusion, 2022, 83–84, 19–52. <https://doi.org/10.1016/j.inffus.2022.03.009>.
- [79] Sano, T., Horii, T., Abe, K., and Nagai, T.: Temperament Estimation of Toddlers from Child–Robot Interaction with Explainable Artificial Intelligence. Advanced Robotics, 2021, 35, 1068–1077. <https://doi.org/10.1080/01691864.2021.1955001>.
- [80] Ring, L., Shi, L., Totzke, K., and Bickmore, T.: Social Support Agents for Older Adults: Longitudinal Affective Computing in the Home. Journal of Multimodal User Interfaces, 2015, 9, 79–88. <https://doi.org/10.1007/s12193-014-0157-0>.
- [81] 小嶋秀樹: 自閉症からみたヒトとロボットのエージェンシー. 日本ロボット学会誌, 2013, 31, 9, 844–845. <https://doi.org/10.7210/jrsj.31.844>.

Appendices

The appendices include the questionnaire sheets used in the experiment and the experimental results of all 13 subjects omitted in Chapter 5. Beginning on the next page, these are presented in order: the pre-experiment questionnaire sheet (original Japanese version and English translation), the post-dialogue questionnaire sheet (original Japanese version and English translation), the table of the post-dialogue questionnaire results, and the dendrograms of Subjects 1–13 from the hierarchical cluster analysis of the post-dialogue questionnaire results.

実験前

2023. . No.

今の気持ち

不快・怒・悲・鬱

快・喜・楽・躁

あなたの気分

--	--	--	--	--	--

あなたについて

嫌い・苦手・難しい

好き・得意・簡単

人と会うのが

--	--	--	--	--	--

人の話を聞くのが

--	--	--	--	--	--

人に向かって話すのが

--	--	--	--	--	--

人と仲良くなるのが

--	--	--	--	--	--

「会話するAI」について

嫌・抵抗がある・無駄

楽しみ・気軽・面白い

AIに向かって話すのが

--	--	--	--	--	--

自由記述

Pre-experiment

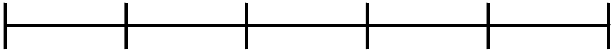
2023. . No.

Current feelings

Negative

Positive

Your mood

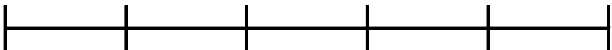


About you

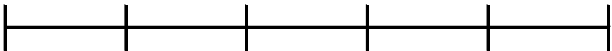
Dislike

Like

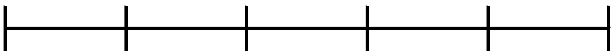
Meeting people



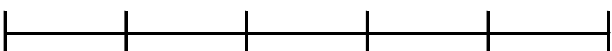
Listening to people



Speaking to people



Becoming friends
with people



About conversational AI

Negative

Positive

Talking to AI



Free comments

実験後(回目)

2023. . No.

この「会話するAI」の印象

そう思わない

そう思う

好意的						
ぎこちない						
自立している						
誠実						
刺激になる						
人格を感じる						
無機質						
話しかけやすい						

この「AIとの会話」について

そう思わない

そう思う

楽しい						
話を聞ってくれる						
スムーズに会話ができる						
別の話題でも話したい						

今の気持ち

不快・怒・悲・鬱

快・喜・楽・躁

あなたの気分						
--------	--	--	--	--	--	--

自由記述 話しながら思ったこと、このAIに言いたいこと等

Post-dialogue(Time:)

2023. . No.

The impression of this conversational AI

	Disagree					Agree
Friendly						
Awkward						
Independent						
Sincere						
Inspiring						
Has personality						
Inhuman						
Easy to talk to						

About this conversation with AI

	Disagree					Agree
Fun						
It listens						
Can have a smooth						
conversation						
Want to talk about						
other topics too						

Current feelings

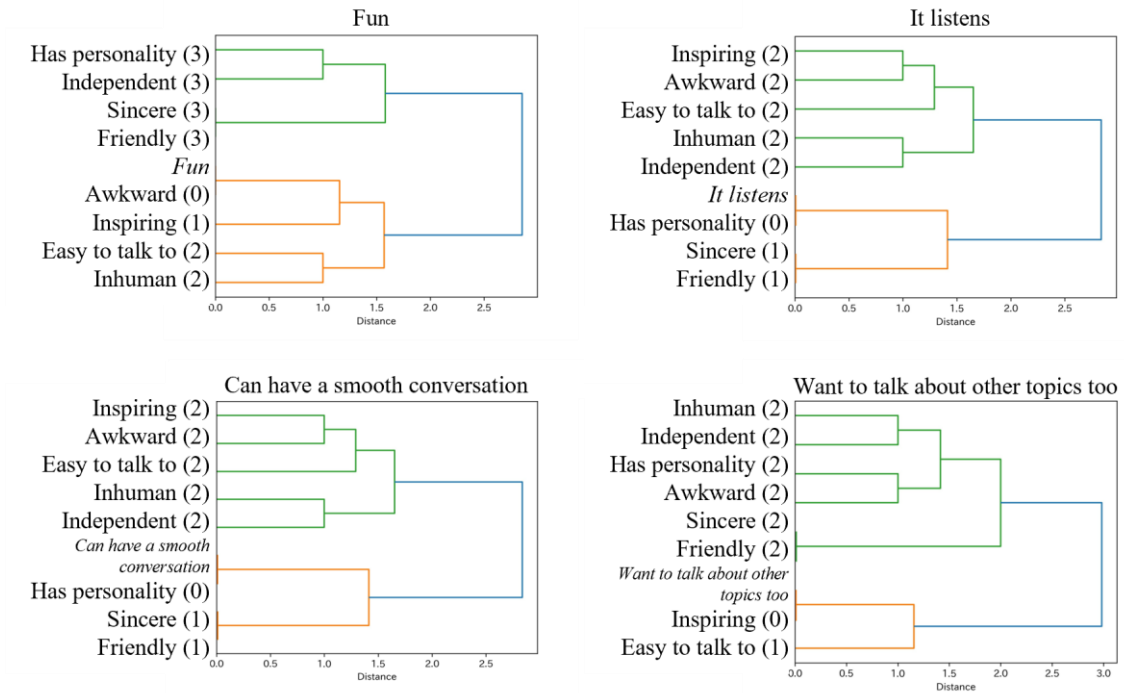
	Negative					Positive
Your mood						

Free comments: thoughts that came to mind while talking, things you want to say to this AI, etc.

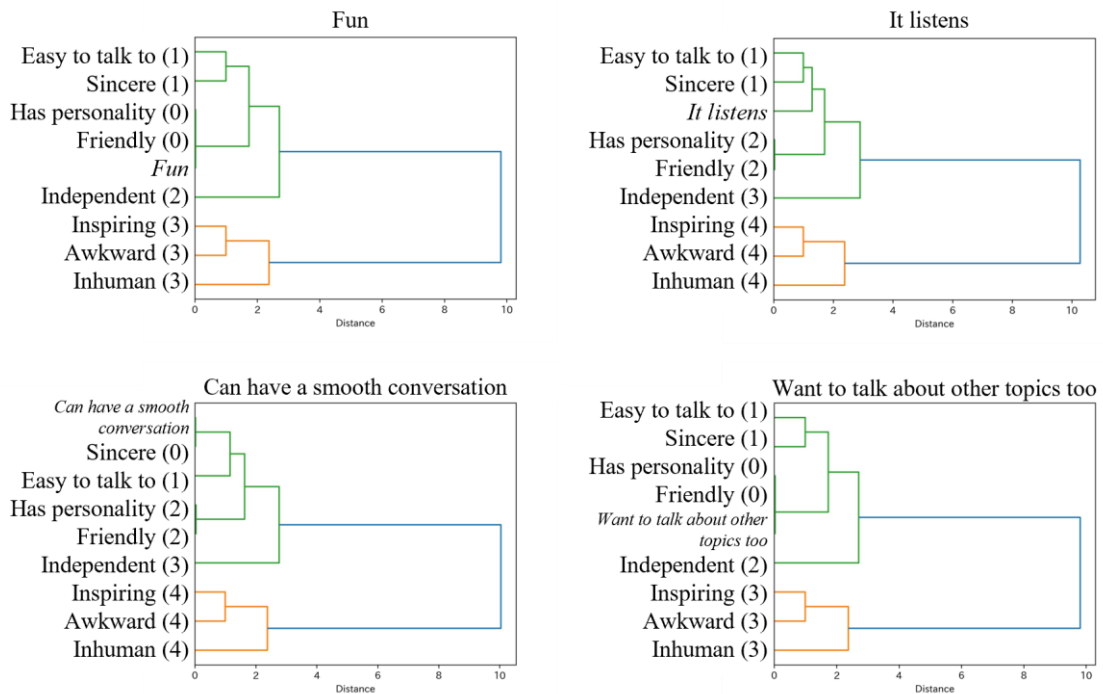
The post-dialogue questionnaire results

Setting	Subject No.	Evaluation Item											
		System						Dialogue					
		Friendly	Awkward	Independent	Sincere	Inspiring	Has personality	Inhuman	Easy to talk to	Fun	It listens	Can have a smooth conversation	Want to talk about other topics too
Dyadic Interaction (with Subject)	1	4	3	3	4	2	3	3	2	3	3	3	2
	2	4	1	4	5	1	4	1	5	4	5	5	4
	3	4	4	4	4	4	4	4	3	4	4	3	4
	4	6	2	5	6	4	5	3	6	5	5	5	5
	5	5	3	3	4	4	3	3	4	5	4	4	4
	6	5	2	5	5	4	3	4	5	5	5	5	4
	7	3	3	4	5	2	3	5	2	2	2	1	4
	8	5	2	6	5	5	4	3	4	4	5	5	5
	9	5	3	4	4	3	4	3	4	4	3	3	4
	10	5	2	4	5	3	4	2	5	6	5	5	5
	11	5	3	4	4	5	5	2	5	4	4	4	4
	12	6	1	6	6	6	6	1	6	6	6	6	6
Dyadic Interaction (with Topic)	13	3	5	4	4	3	2	4	3	3	4	4	3
	1	2	2	3	2	2	2	2	3	2	2	2	2
	2	2	1	4	2	1	2	3	3	2	2	2	2
	3	2	2	3	2	3	3	4	3	3	3	3	3
	4	2	5	1	4	2	1	6	1	1	1	2	1
	5	3	6	5	3	4	3	5	4	4	2	3	4
	6	1	6	1	2	2	1	1	1	2	1	1	1
	7	4	4	4	4	3	3	3	4	3	5	4	4
	8	5	2	6	6	4	5	2	1	3	3	3	5
	9	1	1	2	2	2	2	2	2	2	2	2	1
	10	5	4	5	4	3	5	3	3	5	3	4	6
	11	3	5	4	4	2	2	5	2	2	2	2	2
	12	6	2	5	6	4	6	1	6	6	6	6	6
Triadic Interaction	13	1	5	6	1	3	1	6	1	2	1	1	1
	1	4	3	4	4	3	4	3	3	3	4	4	3
	2	5	1	4	5	2	5	1	5	5	6	5	5
	3	5	1	5	5	5	5	5	5	5	5	5	5
	4	4	4	3	5	3	2	5	3	2	3	4	2
	5	4	4	3	3	4	2	5	4	5	2	3	3
	6	4	6	3	4	2	1	6	2	2	2	2	2
	7	4	3	4	5	3	2	6	2	2	3	3	5
	8	5	2	6	5	5	4	2	5	5	6	5	5
	9	2	5	3	3	2	1	1	2	2	3	2	2
	10	4	5	4	3	3	3	5	4	4	3	2	5
	11	4	3	4	5	5	2	4	5	5	6	4	4
	12	6	1	5	6	5	6	1	6	6	6	6	6
	13	6	1	4	6	5	6	1	5	6	6	6	6

Dendrograms from the hierarchical cluster analysis

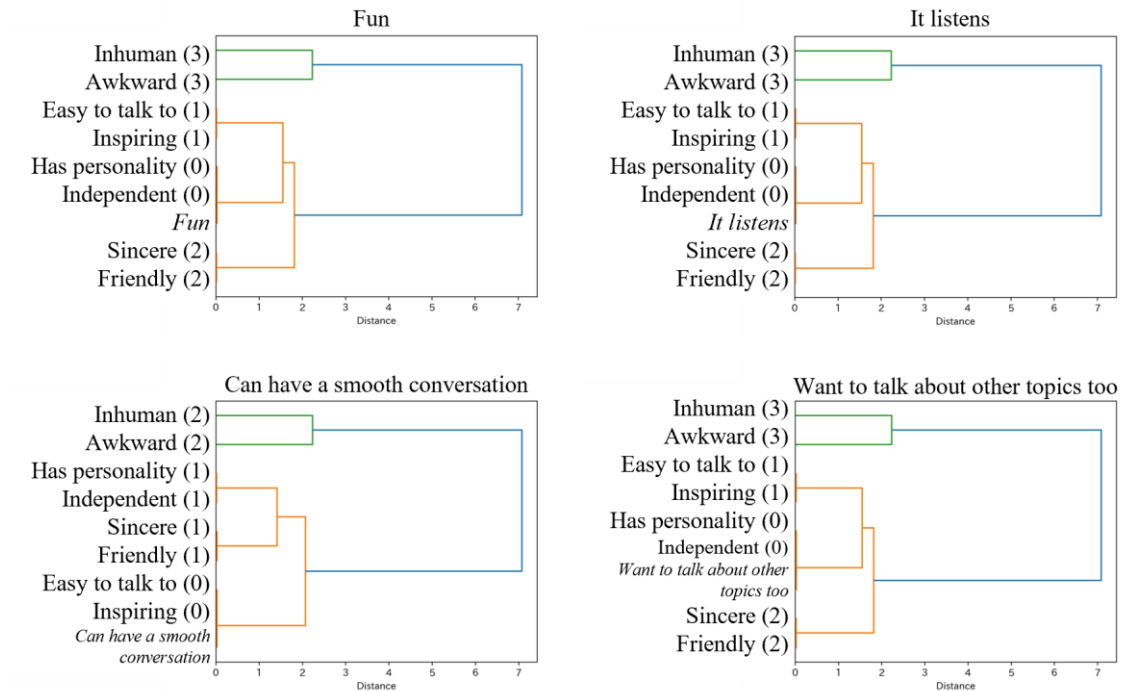


Subject 1

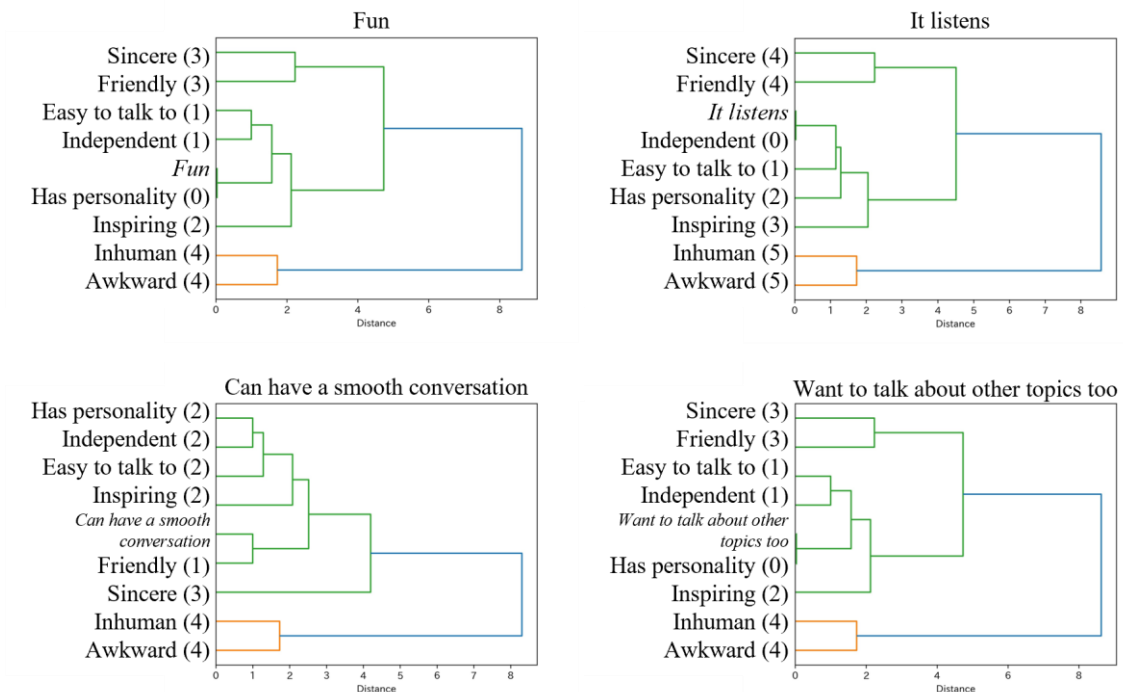


Subject 2

Dendrograms from the hierarchical cluster analysis

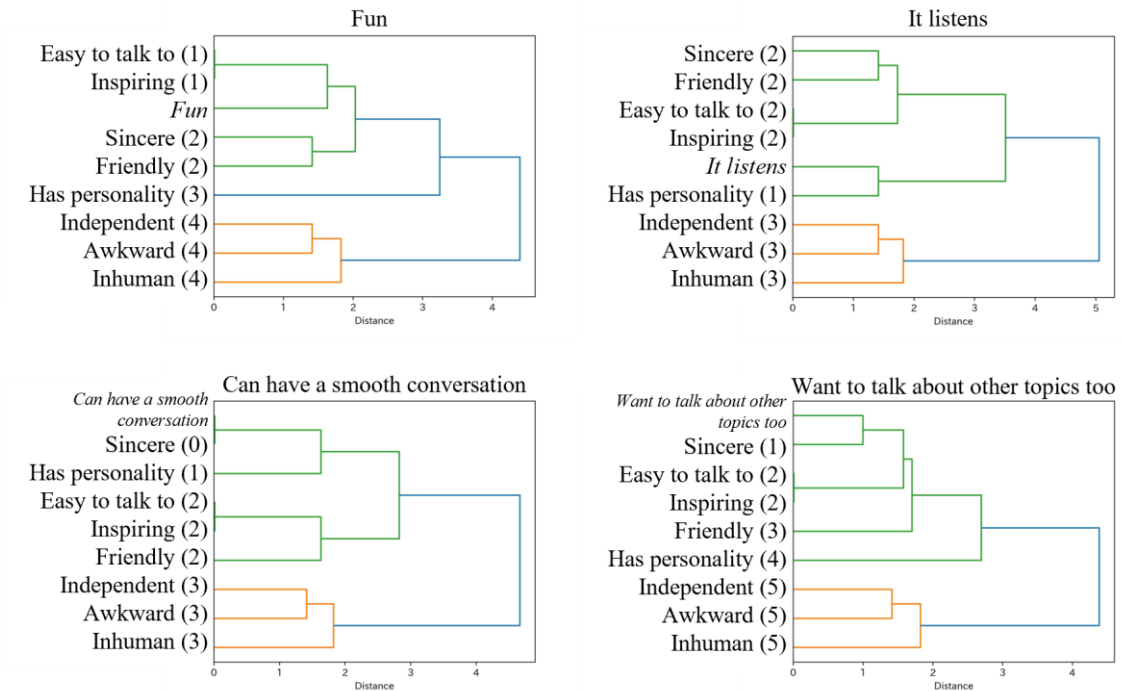


Subject 3

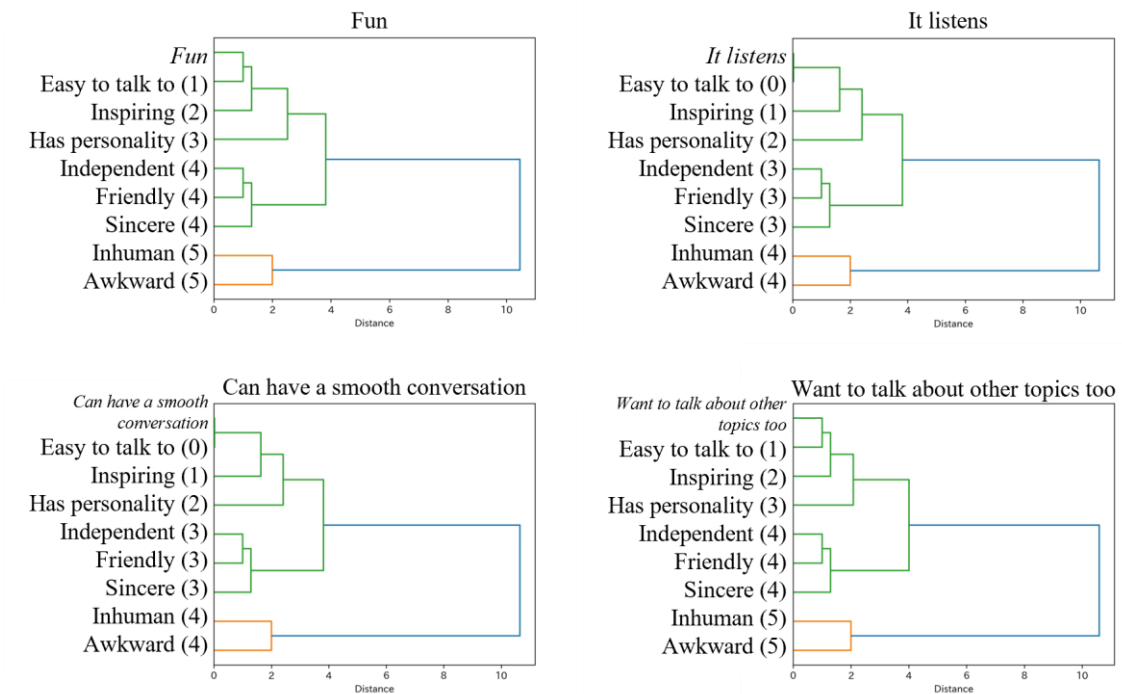


Subject 4

Dendrograms from the hierarchical cluster analysis

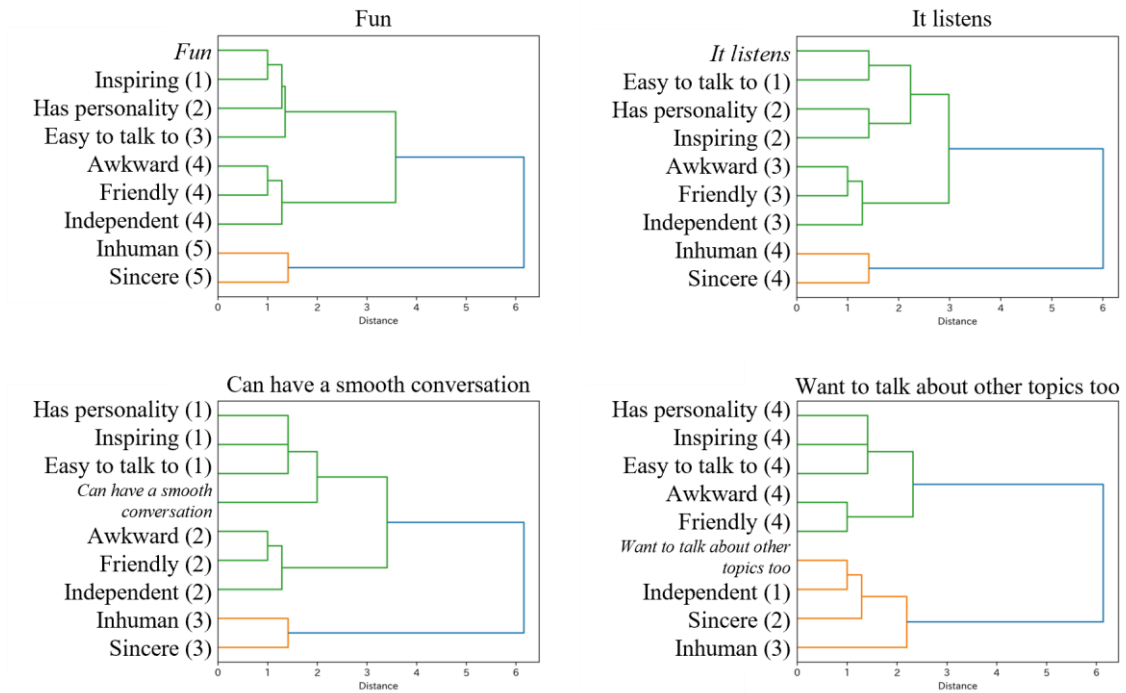


Subject 5

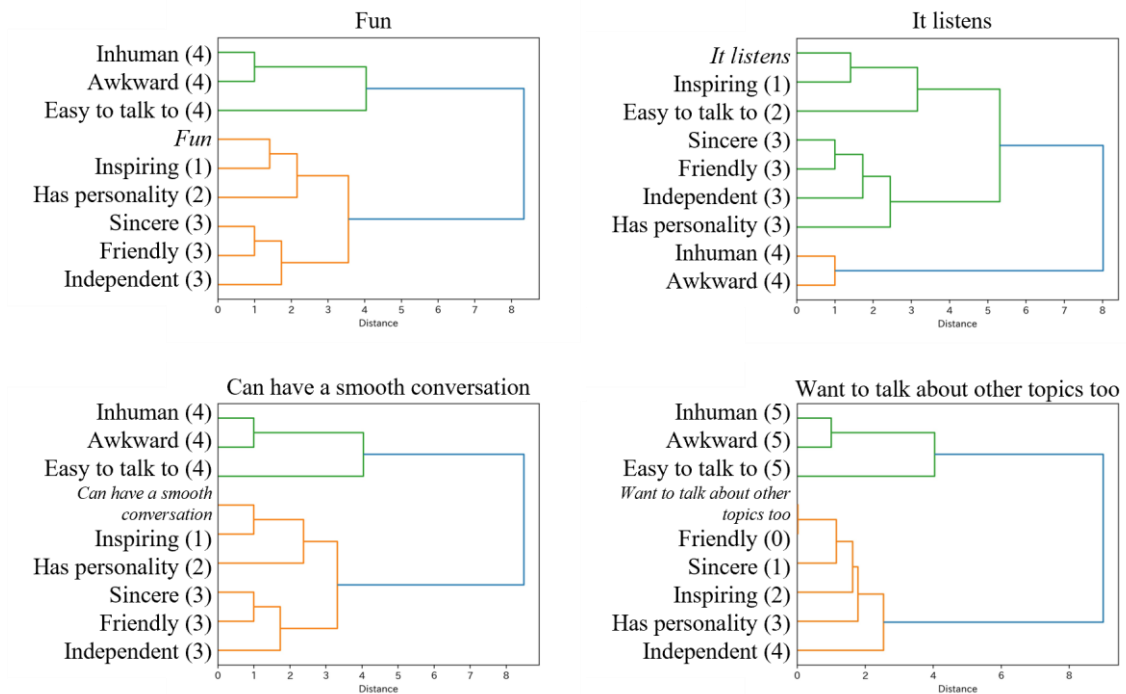


Subject 6

Dendrograms from the hierarchical cluster analysis

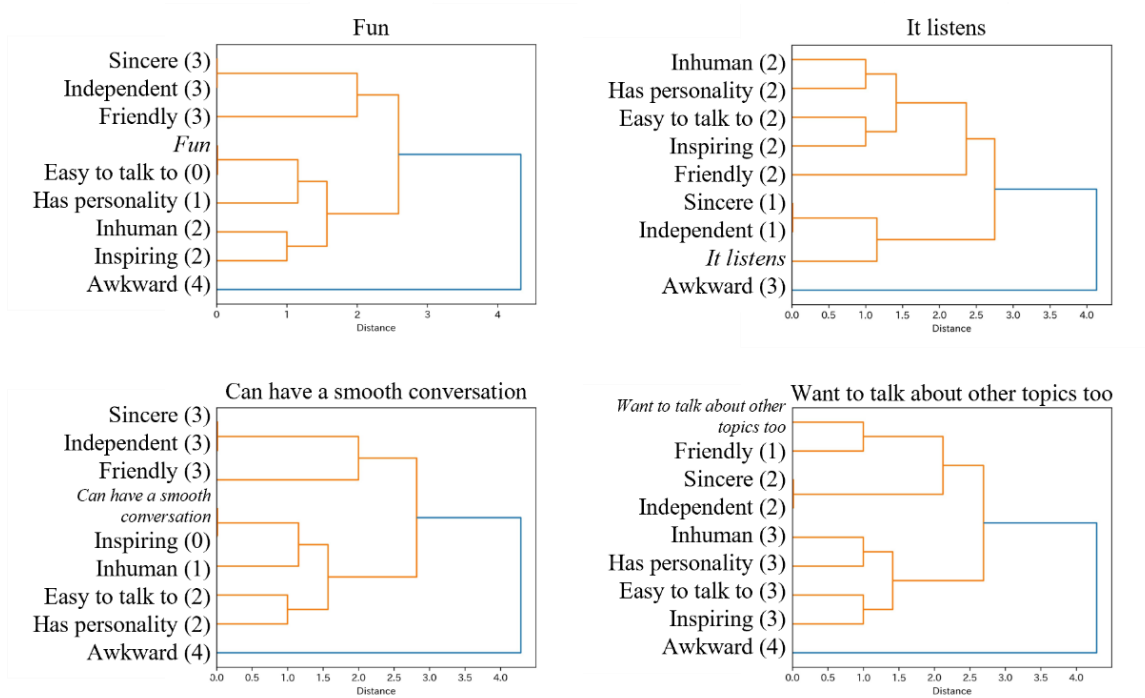


Subject 7

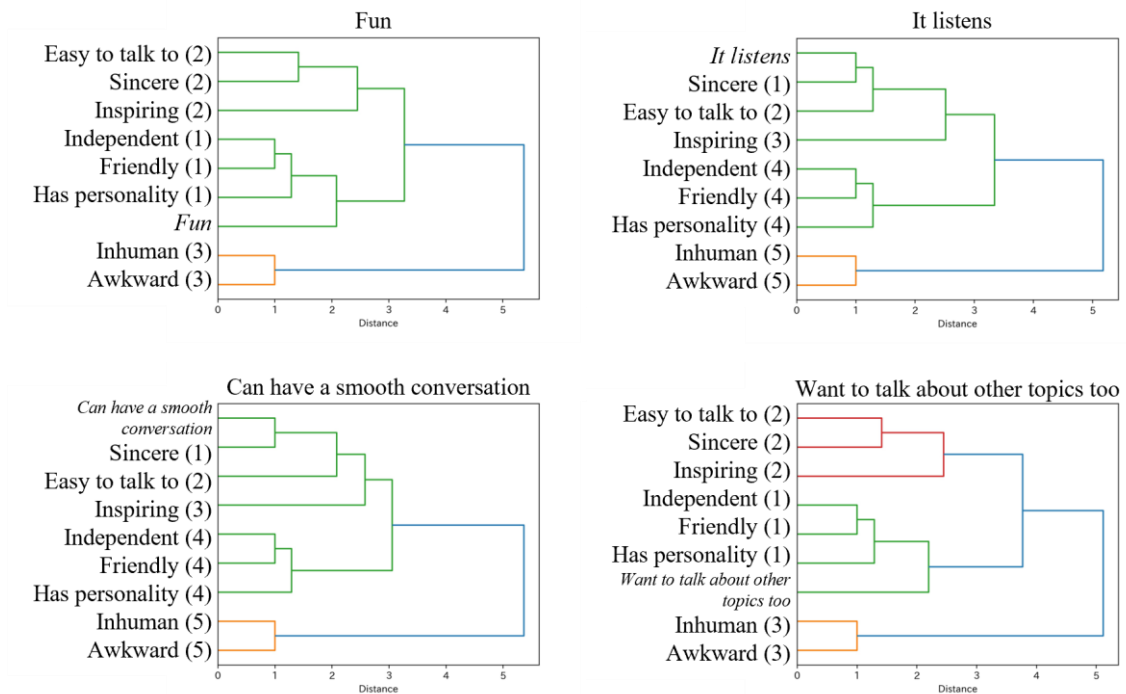


Subject 8

Dendrograms from the hierarchical cluster analysis

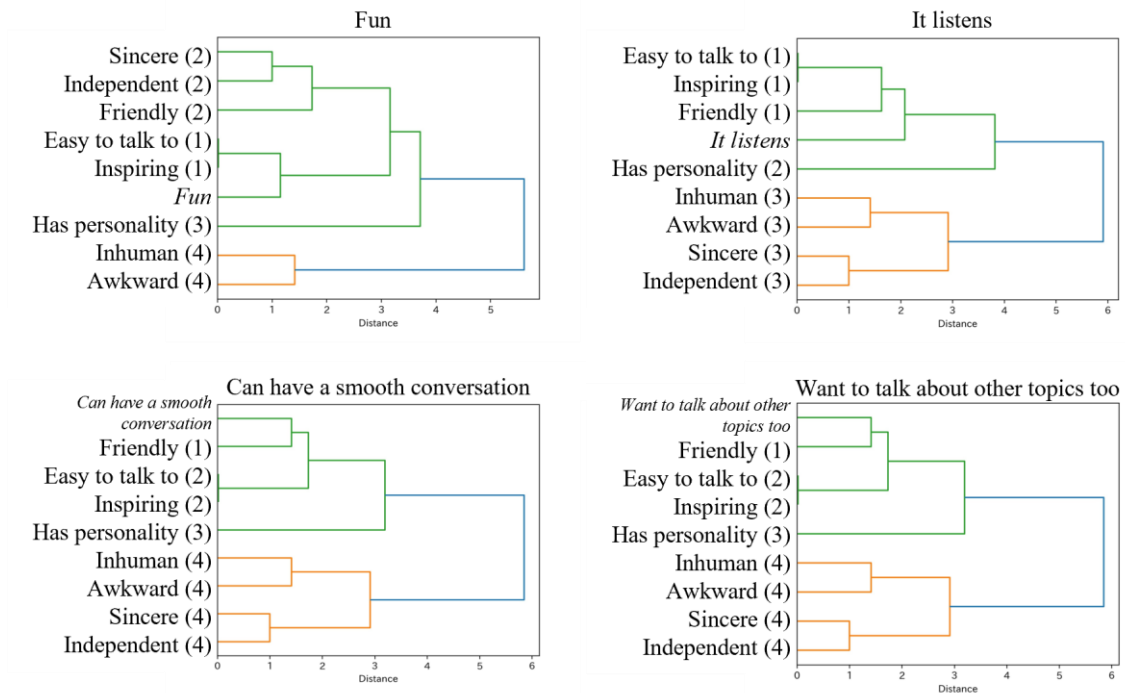


Subject 9

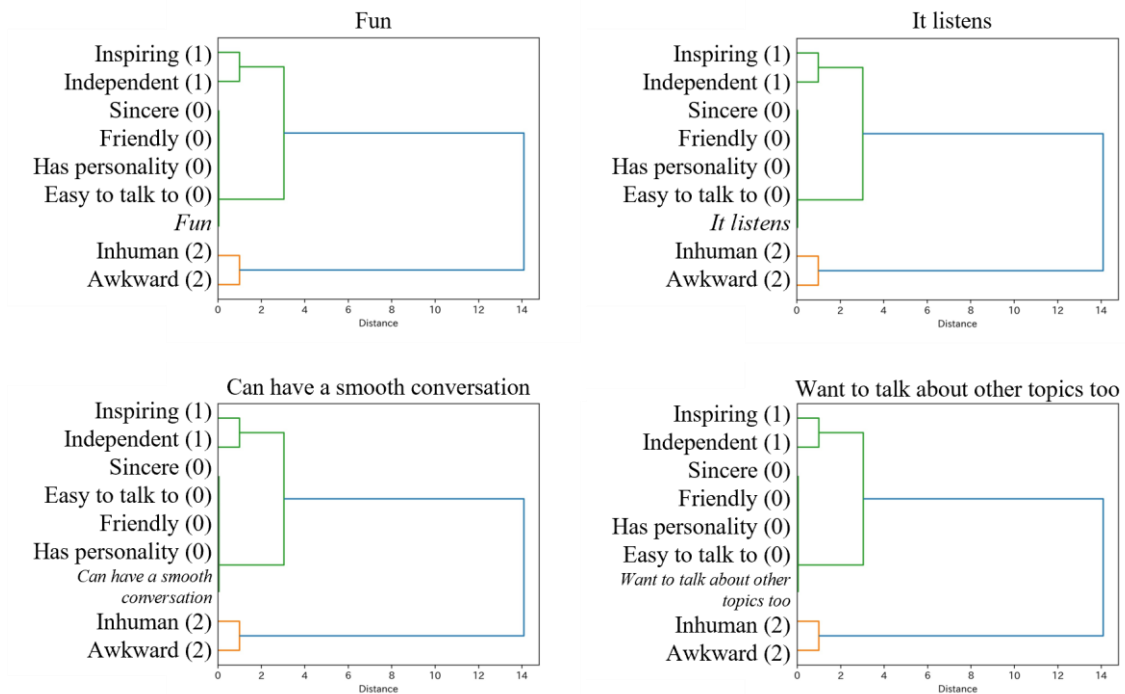


Subject 10

Dendrograms from the hierarchical cluster analysis

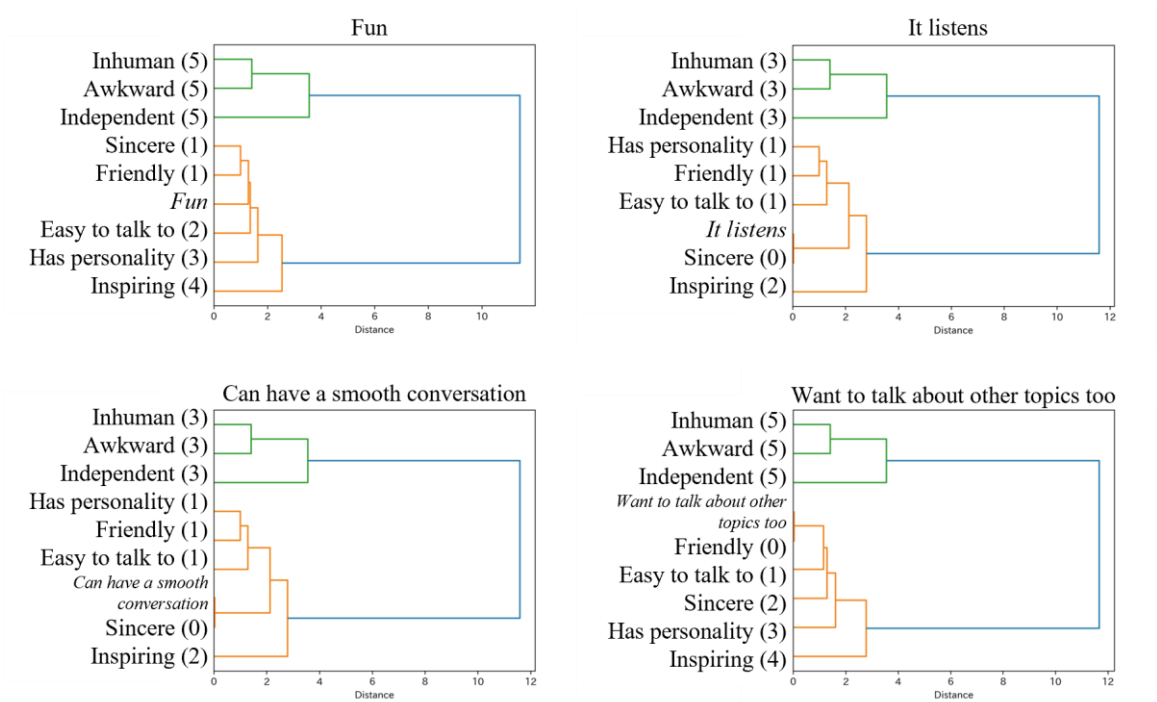


Subject 11



Subject 12

Dendrograms from the hierarchical cluster analysis



Subject 13

Publications

Abe, K., Quan, C., Cao, S., and Luo, Z.: Subjective Evaluation of Dialogues with Dyadic and Triadic Interactions Using a ChatGPT-based Dialogue System. Proceedings of the 2024 Joint 13th International Conference on Soft Computing and Intelligent Systems and 25th International Symposium on Advanced Intelligent Systems (SCIS&ISIS), 2024. <https://doi.org/10.1109/SCISISIS61014.2024.10760216>.

Abe, K., Quan, C., Cao, S., and Luo, Z.: Classification of Properties in Human-like Dialogue Systems Using Generative AI to Adapt to Individual Preferences. Applied Sciences. 2025, 15, 3466. <https://doi.org/10.3390/app15073466>. **[Feature paper]**

Abe, K., Quan, C., Cao, S., and Luo, Z.: Subjective Evaluation of Generative AI-Driven Dialogues in Paired Dyadic and Topic-Sharing Triadic Interaction Structures. Applied Sciences. 2025, 15, 5092. <https://doi.org/10.3390/app15095092>.

Doctor Thesis, Kobe University

“Developmental Scientific Research on Generative AI-based Dialogue System”, 90 pages

Submitted on July 14th, 2025

When published on the Kobe University institutional repository *Kernel*, the publication date shall appear on the cover of the repository version.

© ABE KAORI

All Right Reserved, 2025