Multiparty Conversation Facilitation Robots

February 2015

Yoichi MATSUYAMA

Multiparty Conversation Facilitation Robots

February 2015

Waseda University
Graduate School of Fundamental Science and Engineering,
Major in Computer Science and Engineering,
Research on Perceptual Computing

Yoichi MATSUYAMA

Multiparty Conversation Facilitation Robots

by

Yoichi MATSUYAMA

Submitted to the Major in Computer Science and Engineering
in February 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

In this dissertation, we study a framework for conversational robots facilitating multiparty conversations, which can maintain a group as a group, support group task achievements, and furthermore, entertain a group conversation itself. Starting with reviewing literature about theoretical frameworks of small group dynamics and participation structure, which have been discussed in fields of social psychology, linguistics and cognitive science, we then present a computational model of facilitation processes in multiparty conversations. The process mainly consists of procedural behavior selection regulating socially imbalanced situation and language generation for enjoyable conversations. The facilitation robot plays an unique role observing situations and taking initiatives to regulate equality of engagement density among participants. The procedural behavior production policy is optimized as a partially observable Markov decision process. The results of user studies conducted to evaluate the proposed procedures show evidences of their acceptability of robot's behaviors and feeling of groupness perceived by participants. In the language generation process, we propose an automatic expressive opinion sentence generation mechanisms for enjoyable conversations. Expressed opinions are extracted from a large number of reviews on the web, and ranked in terms of contextual relevance, length of sentences, and amount of information represented by the frequency of adjectives. The sentence generator also has an additional phrasing skill. The results of user studies implied that mechanisms effectively promote interlocutors' enjoyment and interests. As a robotic platform realizing the facilitation model above, we present the SCHEMA system, including its hardware design and network protocols interpreted among software modules. We also present the NANDOKU, a party game system for elderly care as an application of facilitation robots. Then we summarize the dissertation, and discuss future directions of multiparty conversation facilitation robots.

Thesis Committee

Tetsunori Kobayashi	Professor, Faculty of Science and Engineering, Waseda University
Yasuo Matsuyama	Professor, Faculty of Science and Engineering, Waseda University
Tatsuo Nakajima	Professor, Faculty of Science and Engineering, Waseda University
Mikio Nakano	Principal Researcher, Honda Research Institute Japan

Acknowledgments

It was 2005 in Los Angeles, where professor Tetsunori Kobayashi and I met for almost the first time. We were in a party hall for a reception of SIGGRAPH, an international conference of computer graphics and interactive technologies, coincidentally standing next to each other. I had just graduated from the school of literature, majoring in psychology and media studies, dreaming to become sort of a visual media director. He said, "You're a creator, aren't you? When you create something by yourself, you might put your philosophy in it. Research is exact the same thing. We, engineers or scientists, put our philosophies into our products." In my hotel room at that night, I could not sleep overnight, staring at the ceiling and recalling the conversation. A few months later, I would find myself at his office and ask for entering the graduate school to join his research group. "Alright," said the professor, "Now I trust you. It means I'd go around with your adventurous journey. You promise me that you'll show me your own *philosophy* in the near future." Almost ten years passed. I am hereby submitting a product representing my ten-year philosophy to keep the promise.

In the summer of 2007, the very first prototype of multiparty conversation facilitation system was developed. I still remember the moment that ROBISUKE generated the first spontaneous action, "How about you?," addressing me to ask for my opinion. We were excited and showed the professor a video sequence of the interaction. He was staring at the display without saying anything for a while, and finally said he, "Get other faculty members to join us now. This is a very interesting moment." In June of 2008, we brought ROBISUKE to the Care Town Kodaira, an eldery care facility for the first time to show a demo of NANDOKU quiz game playing with elderly people. ROBISUKE's action and utterances did trigger everyone's big laughs. The conversation continued about an hour until the moment when we stopped the system because all the utterance data we prepared was used up. We all were feeling that something socio-psychologically and technologically important was occurring right there. In the end of november of 2009, we finally created our new platform, SCHEMA, which was a successor of previous legendary Waseda robots: ROBITA, ROBISUKE, KOBIAN and HABIAN. Human conversation has been always fascinating and inspiring me. Such unintentional phenomena in our daily lives has deep science in it. I have been exploring it day and night with these robots.

Each moment was unforgettable. I could never achieve them alone. There are so many people who have given me great advises, contributed ideas through discussions, and hacked codes for the projects. First of all, I would like to thank again professor Tetsunori Kobayashi, who has given me a chance to start this grand journey. His words always led me in the right directions. Also, I would like to thank the thesis committee: professor Yasuo Matsuyama, professor Tatsuo Nakajima and professor Mikio Nakano for the long and tough review and judgement process. Their comments and suggestions have been so very helpful and invaluable. Professor Machiko Kusahara, an advisor of my undergraduate degree, as well as a great curator, media artist and educator, inspired my future carrier combining arts and sciences.

I cannot thank enough Shinya Fujie, a great leader and collaborator throughout projects. I could not achieve anything without his insights, knowledge and experiences. Tetsuji Ogawa and Teppei Nakano have gave me their personal and professional advices. Shinsuke Akaike developed an early facilitation system on ROBISUKE in 2007. Hikaru Taniyama largely contributed the SCHEMA platform development in 2009. The initial version of the SCHEMA QA was developed by a collaboration with Yushi Xu, a visiting researcher from MIT CSAIL in 2010, and its automatic sentence generation system was developed by a collaboration with Akihiro Saito. Iwao Akiba helped me a lot for modeling and developing the facilitation strategy. There are many other contributors at the Perceptual Computing Group: Hiroki Tsuboi, Kosuke Hosoya, Atsushi Ito, Yusuke Kinoshita, Masanori Kikuchi, Esuke Soma, Moemi Watabe, Azusa Todoroki,

Kenshiro Ueda, Tomoki Hayashida, and many other students. And the field experiments were able to be conducted thanks to Kaoru Nishikiori, a care manager of the Care Town Kodaira.

Next, I would like to thank mentors worldwide. Professor Justine Cassell at the Articulab, the Human-Computer Interaction Institute, Carnegie Mellon University hosted my visit in 2014. The collaboration with her and the ArticuLab team was so much exciting and was a great opportunity to think about my own research in larger contexts. Professor Giorgio Metta at the iCub Facility, Italian Institute of Technology hosted my visit in 2013. I believe iCub is one of the best cognitive robot platform. I was so fascinated by the sophisticated system created by Europe-wide excellent research communities. Professor Carolyn Rosé at the Language Technology Institute, Carnegie Mellon University mentored my SIGDIAL paper.

I also would like to thank co-founders of WIZDOM (Waseda University Integrated Space of Wizards, Digital Oriented Manufacturers): Akihiro Hayashi, Atsushi Enta, Jun Nakagawa, Kosuke Kikuchi and Taiki Watai. The multi-disciplinary activities with them fueled my energies of creation.

This research has been supported by the following grants: Grant-in-Aid for scientific research WAKATE-B (23700239) "Development and Evaluations of Multiparty Conversation Activation Systems (2010-2012)," WAKATE-B (25870824) "Facilitation Strategy for Multiparty Conversation Robots (2013-2015)," International Research and Education Center for Ambient Soc, Waseda University Global COE Program (2008-2010), Microsoft Scholarship (2009), and Japan Society for the Promotion of Science (JSPS) Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation (2013-2015). TOSHIBA corporation provided the speech synthesizer customized for our spoken dialogue system.

Finally, I thank my family: my father Toshiro, my mother Tomiko and my sister Tomoko, who have all stood by me through the best and the worst times. And I especially would like to dedicate this thesis and doctoral degree to my grandfather Shoji who passed away in 2008, the biggest supporter in my life. He should have been very glad at my accomplishment. My uncle Shoichiro Saito also has been another big supporter.

This is a destination of my youthful journey and a beginning of new one. My wife Hiroko has been always encouraging me to enjoy the whole process. We now begin a new journey far beyond today.

Contents

List of Figures

List of Tables

"They say that knowledge born of experience is mechanical, but that knowledge born and consummated in the mind is scientific, while knowledge born of science and culminating in manual work is semi-mechanical. But to me it seems that all sciences are vain and full of errors that are not born of experience, mother of all certainty, and that are not tested by experience, that is to say, that do not at their origin, middle, or end pass through any of the five senses."

Leonardo da Vinci

1 Introduction

1.1 Background and Purpose of the Dissertation

Conversation is a core natural function of human beings, which enables them to interact with others and organize societies. The recent advancement of social neuroscience indicates that human brain is mostly optimized for social activities (Gazzaniga, 1985; Brothers et al., 2002), and it has been evolved to survive social environments (Dunbar, 1998). So, what is the best way to understand such social functions? Leonardo da Vinci emphasized the importance of the combination of scientific and engineering endeavors to understand nature, meaning that a complete understand of nature requires a whole process from an analysis of nature to a creation of a synthesized product representing a certain aspect of nature. We follow the Leonardo's way that might require a certain level of abstraction. Here, we attempt to abstract the social phenomena in terms of conversational protocols exchanged among humans through their bodies. Embodiment plays an important role in this context because embodiment functionally equivalent to a human body is a crucial factor to interpret and synthesize the conversational protocols. Such a way would allow realizing a "conversationally smart" agent (Cassell, 2000). Furthermore, our social environment doesn't always exist as a pair of persons, but sometimes as a group. As Dunbar, an anthropologist, argued that human brain evolved to survive and reproduce in large and complex social groups (Dunbar, 1998), social intelligence most likely appears in the nature of group. While traditional research of embodied conversational agents have focused on a dyadic situation, it should have expandability to a group model.

Human talks with human. From our birth to death, we all talk with others everyday. Such daily phenomena are eagerly desired by us for well-being. Today, significant numbers of advanced countries are facing serious population ageing. According to the Annual Report on the Aging Society in Japan (Cabinet Office, 2014), the population aging rate (over 65 years old) in 2013 was up to 25.1%, and pursuing quality of life (QOL) has been a major policy goal. The demographic change (rapid population aging resulting from the decline in the birth rate) itself is not negative phenomenon, but as Peter Drucker pointed out, it is one of opportunities of innovation (Drucker, 1984). We have investigated continuously at one of daycare centers

in a suburb of Tokyo. As the result, we realized that communication is desired for its own in such facilities and communicating with other people could cure even depression and dementia. Conversation with other person has been a killer application of human cognition. So, why not creating a robot to *support* human-human activities, not to *replace* a human interlocutor by them if they were capable enough? In fact, active communication among elders or between an elder and care staffs can not be always spontaneously emerged in such care facilities because of lack the manpower to keep watch statuses of all residents. A robot situated in conversational situations with its embodiment, and has capabilities to understand and generate the human conversation protocols, has a big potential to mediate human-human conversations fundamentally desired by human nature.

Such physically situated conversational robots have two major technological backgrounds: spoken dialogue systems and human-robot interaction. Ichiro Kato, a roboticist at Waseda University originated the first full-scale anthropomorphic humanoid typed robot in the world in the early 70's, which was named WABOT-1. WABOT-1 was an epoch not only because it could walk like human, but also it has speech interaction capabilities, including speech recognition and synthesis (Kato, 1973). Since then, Waseda humanoid group have developed numerous humanoid robots. WABOT-2 is an humanoid robot playing keyboard instruments (Kato et al., 1987). It could understand musical score and speech command via continuous speech recognition. Such robots were integrated systems combining their physical embodiment and conversational capabilities. However, these components have been diverged into robotics and speech technologies and seldom been integrated. In order to develop natural human-machine interfaces, numerous spoken dialogue systems have been developed. ELIZA, one of the origin of dialogue systems created by Joseph Weizenbaum, could receive natural language text and reply using primitive pattern matching techniques to simulate a psychotherapist (Weizenbaum, 1966). Although it only has a surface level of language processing without semantic processing, many people was addicted to talk with ELIZA. Most of the current dialogue systems are successors of ELIZA in terms of language processing. VOYAGER is one of the early examples of sophisticated spoken dialogue system (Glass et al., 1995). Differently from text-based dialogue, spoken dialogue has many difficulties in terms of conversational protocols shared with human. The recent trend of dialogue systems takes advantages of statistical learning. Let's Go system is a large scaled spoken dialogue system providing bus information for the general public including the elderly and non-natives (Raux et al., 2005). Young et al. proposed a partially observable Markov decision process (POMDP) based spoken dialogue system (Young et al., 2010). Recently, numerous industrial spoken dialogue systems have been released. Apple Inc.'s Siri[1] is a personal assistant and knowledge navigator run on iOS mobile devices, which was originated as DARPA's CALO project[2]. Microsoft Corp. also released Cortana[3] on Windows Phones. GoogleNow[4] is a proactive personalized information recommendation system. NTT Docomo's Shabette Concier[5] is also a personal assistant system that has question answering capabilities (Uchida et al., 2013). In the field of Human-Robot Interaction, an intersection of robotics, artificial intelligence, psychology, social science and design, nonverbal aspect of interactions, including eye gaze, facial expressions and interpersonal distance, have largely been investigated. In these studies, robots' embodiment shared with human is supposed to be a critical factor of natural interaction. In 2010, a small symposium named "Dialog with Robot[6]" was held, which was an attempt to reunite robotic research and dialogue research fields again (Bohus et al., 2011).

[1] https://www.apple.com/ios/siri/
[2] http://www.ai.sri.com/project/CALO
[3] http://www.windowsphone.com/en-us/how-to/wp8/cortana/meet-cortana
[4] http://www.google.com/landing/now/
[5] https://www.nttdocomo.co.jp/service/information/shabette_concier/
[6] http://hci.cs.wisc.edu/aaai10/

Fuleled by such a scientific desire and, social and technological backgrounds, we explore design of a facilitation robot system in multiparty conversational setting to support human-human conversations, which could unite research of physically situated robot systems and multimodal dialogue mechanisms. Specifically, in this dissertation, we formalize facilitation strategies and multimodal language generation process. We then consider a robotic platform to implement these mechanisms. Also, based on the fundamental considerations, as an application of multiparty conversation facilitation robots, we attempt to develop a robot that can participate in a party game taking place in a elderly care facility to entertain other participants.

1.2 Related Work

In this section, we review related work attempting to develop embodied social agent systems. Starting with dyadic model, we then turn to multiparty model.

1.2.1 Embodied Conversational Agents for Dyadic Interactions

Numerous research on two-participant interaction models enabling sociable and conversational capabilities have been conducted. Cassell et al. pioneered Embodied Conversational Agents (ECA) (Cassell, 2000). They argued that embodied conversational agents should be "conversational smart," which has social and linguistic intelligence for face-to-face conversation. They presented an ECA architecture fulfilling some requirements, including the use of multimodal input and output, and the ability to deal with conversational interaction mechanisms such as turn-taking and feedback. Breazeal presented a sociable robot, Kismet, that can engage with humans in expressive social interaction (Breazeal, 2004). Although Kismet does not have linguistic capabilities, therefore it is not a conversational agent in this context, Kismet perceives a variety of natural social cues from visual and auditory channels, and delivers social signals to people through gaze direction, facial expression, body posture, and vocalization. These social competencies were inspired by infant social development, psychology, ethology, and evolutionary perspectives. Thomaz studied use of human guidance for machine learning of sociable robots using a reinforcement learning framework, namely the Socially Guided Machine Learning, which can be successfully incorporated with improving learning performance (Thomaz et al., 2006). With respect to conversational functions, Raux et al. presented the Finite-State Turn-Taking Machine (FSTTM), a model to control the turn-taking behavior of conversational agents (Raux and Eskenazi, 2009). Chao et al. presented a computational model and architecture for situated turn-taking, namely, the CADENCE (Control Architecture for the Dynamics of Embodied Natural Coordination and Engagement) using their robotic platform, Simon (Chao and Thomaz, 2012b,a; Thomaz and Chao, 2011). Fujie et al. developed a conversation robot with back-channel feedback function based on linguistic and nonlinguistic information (Fujie et al., 2004). In order to determine the content of the feedback earlier than the end of the utterance, they used finite state transducer based speech recognizer that outputs the content of the feedback. And they used prosody information, especially the fundamental frequency (F0) and the power of the utterance, to extract the proper timing of the feedback. Rich et al. developed an computational model for recognizing engagement between a human and a humanoid robot based on a study of the engagement process between humans (Rich et al., 2010).

1.2.2 Multiparty Conversational Agents

In the context of agents for multiparty conversation in virtual worlds, Traum et al. developed a conversational agents interacting with multiple users using verbal and non-verbal modalities (Traum and Rickel,

2002). Zheng et al. developed Elva, an embodied tour guide that facilitates multi-party interaction in an interactive art gallery environment (Zheng et al., 2005). The agent acted as a leader coordinating and facilitating multiparty interaction in a casual social group, specifically, a tour group context. Rudnicky et al. developed TeamTalk, a platform for multi-human-robot spoken dialog systems in coherent real and virtual spaces in order to explore research free from mundane allocation constraints and speed-up our platform development cycle (Rudnicky et al., 2010). Merge et al. presented a method for situated agents to recover from miscommunication using the TeamTalk platform (Marge and Rudnicky, 2011).

In the context of physically situated agent, Matsusaka et al. pioneered multiparty conversational robots (Matsusaka et al., 2003). They developed a conversational robot fulfilling requirements to participate in a group conversation, including understanding status of participation, and ways of transferring messages to other participants. They assumed participants can be divided into two groups, one is "parties concerned" and the other is "observers". In the parties concerned group, a participant who has a turn is the "speaker (primary receiver)", and participants belonging to the observers group are secondary receivers. Isbister et al. developed CoBot, a helper Agent, which introduces safe topics of discussion to improve group process in a multi-cultural human-human interaction environment (Isbister et al., 2000). It manages and shares user information among a group where users may be looking for their friends. Mutlu, et al. examined the role of eye gaze to establish the participation structure in a conversational group using a Wizard of Oz method with naive subjects (Mutlu et al., 2009, 2012).

Major differences of phenomena between dyadic and multiparty are engagement and addressing (role assignment). Sidner et al. developed an agent system that can engage with users, where they defined engagement as "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake" (Sidner et al., 2004). Bohus et al. modeled engagement in multiparty conversations based on the Sinder's definition, i.e., open world dialogue[7]. In their model, an agent manages the engagement process, including the following components. (1) Sensing the engagement state, actions, and intentions of multiple agents in the scene; (2) making engagement decisions (i.e., whom to engage with, and when); and (3) rendering these decisions in a set of coordinated low-level behaviors in an embodied conversational agent. They evaluated the effectiveness of multimodalities, including gaze, gesture, and speech, for a multiparty conversation facilitating agent (Bohus and Horvitz, 2009a,b,c,d, 2010a,b, 2011a,b). Foster et al. presented another similar interactive kiosk system that handles situated, open-world, multimodal dialogue in scenarios, namely the JAMES, Joint Action for Multimodal Embodied Social Systems [8]. It extends Bohus's Open-World Dialogue system by adding physical embodiment, which has been shown to have a large effect on social interaction. The specific demonstration of this work is a bartender scenario, where the robot plays the role of a bartender responding to customers' requests in a dynamic setting, with multiple customers and short interactions. Interactions in the target scenario will incorporate a mixture of task-based behaviors (e.g., ordering and paying for drinks) and social behaviors (e.g., engaging in social conversation, managing multiple simultaneous interactions), both of which present challenges for the JAMES project: a robot existing in the physical world must be able to understand and respond to both the social and the task-based needs of the humans that it encounters, and to successfully distinguish them from each other (Foster et al., 2012).

Bohus et al anchored their discussion of challenges for open-world dialog in Clark's model of language interaction (Clark and Schaefer, 1989). With this model, natural language interaction is viewed as a joint activity in which participants i n a conversation attend to each other and coordinate their actions on several different levels to establish and maintain mutual ground. Components of Clark's perspective includes (1)

[7] http://research.microsoft.com/en-us/um/people/dbohus/research_situated_interaction.html
[8] http://james-project.eu/

Channel level, (2) Signal level, (3) Intention level and (4) Conversation level. Bohus et al. closely studied the Channel level to managing engagements.

In the upper conversational level in the Clark's model above, interactions with users needs conversational strategies and semantic processes. In terms of facilitation strategies, Kumar et al. have developed and evaluated automated text-based tutors for two different educational domains equipped with eleven social interaction strategies based on Bales' Socio-Emotional Interaction Categories, especially corresponding to three positive Socio-Emotional Interaction Categories: Showing Solidarity, Showing Tension Release and Agreeing (Kumar et al., 2010; Kumar and Rosé, 2010a,b; Kumar et al., 2011; Kumar and Rosé, 2011). Dosaka et al. developed a thought-evoking dialogue system for multiparty conversations with a quiz-game task (Dohsaka et al., 2009). They reported that the existence of agents and empathic expressions is effective for user satisfaction and can increase the number of user utterances. We have previously developed a multiparty quiz-game-type facilitation system for elderly care (Matsuyama et al., 2008) and reported the effectiveness of the existence of a robot (Matsuyama et al., 2010).

1.3 Research Objectives

Based on the background and related work, in this dissertation, the following issues will be taken into account to develop a multiparty facilitation robot system.

Computational Architecture for Facilitation Process

In order to build computational architecture of facilitation process, we will begin with reviewing literatures about theoretical frameworks of small group dynamics and participation structure, which have been discussed in fields of social psychology, linguistics and cognitive science, we then present a computational model of facilitation process in multiparty conversations. Major differences between dyadic and multiparty conversations are mechanisms of engagement and addressing (recipient design and role assignment). In dyadic conversations, an agent's interlocutor is clearly understood by each other. In multiparty conversations, each participant mutually recognizes boundaries of a group. Current speaker designs its utterance to address primary listeners where sometimes agreements of addressed participants are not mutually shared. These complicated mechanisms have been discussed as participation structure.

Engagement Density Control for Facilitation

In multiparty conversations, socially imbalanced situations frequently emerge, where a dominant participants are leading the conversation and the other participants tends to be left behind. In such situations, a facilitator could play an unique role observing situations and taking initiatives to regulate equality of engagement density among participants. In order to regulate the socially imbalanced situations, the facilitator should take procedural steps to obtain initiatives. We will discuss a computational model of procedural behavior generation.

Language Generation Process for Enjoyment

In terms of functional conversations, Grice's Maxim of Quantity suggests that responses should contain no more information than was explicitly asked for. However, in our daily conversations, more informative response skills are usually employed in order to hold enjoyable conversations with interlocutors. We attempt to model language generation process for enjoyable conversations.

Specifications for Physically Situated Robotic Platform

In order to implement a conversational agent system, a robotic platform needs to have capabilities to exchange conversational protocols. Such protocols essentially need robot's embodied functions, including facial expressions, head gestures, and directional control of torso. As for internal protocols transferred among software modules, we review and discuss existing attempted networking middlewares and higher conversational protocols, which allow us to easily develop conversational robot systems.

Practical and Promising Applications of Facilitation Robots

Based on the background of serious situations of aging society in Japan, we will propose an elderly care application as practical and promising applications of facilitation robots. Beginning with observing actual elderly care services at an elderly care facility in Tokyo, we will propose a system design of a party game application entertaining elderly people. We then discuss future direction of facilitation robots based on discussions of the results of field and laboratory experiments.

1.4 Dissertation Organization

In this dissertation, we will study a framework of conversational robots that facilitates multiparty conversations. We will specifically present facilitation strategies to maintain conversational groups, expressive sentence generation methods for enjoyment, design of a robotic platform for multiparty conversation, as well as a promising application of facilitation robots.

In **chapter 2**, in order to study general facilitation framework, we will begin with pointing out differences between dyadic and multiparty conversations, with employing the theories of participation structure. As theoretical frameworks of strategic facilitation processes, research of small group dynamics will be reviewed. We will then attempt the whole architecture for facilitation robots.

In **chapter 3**, we will present facilitation strategies that regulate imbalanced engagement density in four-participant conversation as the forth participant with proper procedures for obtaining initiatives. Four is the spacial number in multiparty conversations. In three-participant conversations, the minimum unit for multiparty conversations, social imbalance, in which a participant is left behind in the current conversation, sometimes occurs. In such scenarios, a conversational robot has the potential to objectively observe and control situations as the fourth participant. Consequently, we will present model procedures for obtaining conversational initiatives in incremental steps to harmonize such four-participant conversations. During the procedures, a facilitator must be aware of both the presence of dominant participants leading the current conversation and the status of any participant that is left behind. We will model and optimize these situations and procedures as a partially observable Markov decision process (POMDP), which is suitable for real-world sequential decision processes. The results of experiments conducted to evaluate the proposed procedures show evidence of their acceptability and feeling of groupness.

In **chapter 4**, we will present the SCHEMA QA, an enjoyable question answering framework that has expressive opinion generation mechanisms. These responses are usually produced as forms of one's additional opinions, which usually contain their original viewpoints as well as novel means of expression, rather than simple and common responses characteristic of the general public. In this chapter, we will propose an enjoyable question answering framework comprising an automatic expressive opinion generator. The opinions generated are extracted from a large number of reviews on the web, and ranked in terms of contextual relevance, length of sentences, and amount of information represented by the frequency of

adjectives. The framework also has additional phrasing skills that enable it to generate small talk. The results of two experiments conducted to evaluate users' enjoyment of the opinion generation and additional phrasing mechanisms indicate that both mechanisms effectively promote users' enjoyment and interests.

In **chapter 5**, we will present the design of SCHEMA robot as a robotic platform for multiparty conversation facilitation. In order to participate in multiparty conversational situations, and be recognized as a ratified participant, a robot needs to have capabilities to exchange conversational protocols, which include organizing participation structure, transmitting messages, and turn-taking. Such protocols essentially need a robot's embodied functions, including facial expressions, head gestures, and directional control of torso. Based on our studies, SCHEMA has 22 degrees of freedom. It was also designed with a user-friendly styling for all generations, from children to elderly people.

In **chapter 6**, we will present a party-game system for elderly care as one of tasks of facilitation robot systems. The robot participates in a quiz game with other participants and tries to activate the game. We will report field experiments in an adult day-care center, and laboratory experiments to evaluate the effectiveness of the robot's behaviors.

In **chapter 7**, we summarize our works and discuss future directions.

"The question for me is, how can the human mind occur in the physical universe? We now know that the world is governed by physics. We now understand the way biology nestles comfortably within that. The issue is, how will the mind do that as well? The answer must have the details. I have got to know how the gears clank and how the pistons go and all the rest of that detail. My question leads me down to worry about the architecture."

Allen Newell

2 Facilitation Framework

2.1 Introduction

In this chapter, in order to describe a facilitation framework, we begin with reviewing literatures about existing general theoretical frameworks of dyadic conversational agents, then extend it to a multiparty model with concepts of small group dynamics and participation structure, which have been discussed in fields of social psychology, linguistics and cognitive science. We then present the SCHEMA Framework, a computational architecture of facilitation processes, which is modeled in terms of conversational protocols.

2.2 Framework for Dyadic Conversation

Traditionally, research on conversation has strong assumption to study dyadic interactions.In general, sociology defines "dyad" as a group of two people, the smallest possible social group. Cassell et al. pioneered investigating frameworks of embodied conversational agents (ECAs) for dyadic interactions (Cassell, 2000; Cassell et al., 1999; Cassell and Bickmore, 2003; Bickmore and Cassell, 1999). They defined ECAs as agents that have the same properties as humans in face-to-face conversations including the following: (1) abilities to recognize/respond to verbal and non-verbal input, (2) generating verbal and non-verbal output, (3) dealing with conversational functions such as turn taking, feedback, and repair mechanisms, and (4) giving signals that indicate the state of the conversation, as well as to contribute new propositions to the discourse. As architectural requirements for ECAs, they concluded that the construction of an agent character that can effectively participate in face-to-face conversation as described above requires the following features:

1. *Multimodal Input and Output*: Since participants in a dyadic conversation send and receive information through gesture, intonation, gaze and speech, an architecture should have capbilities receiving and transmitting such information.

2. *Real-time*: Different threads of communication have different requirements of response timescale (e.g. feedback and interruption occur in milliseconds, while quiestion-answer interactions sometimes occur in seconds).

3. *Understanding and Synthesis of Propositional and Interactional Information*: In order to deal with propositional information with a model of user's needs and knowledge, an architecture should have both a static domain knowledge and a dynamic discourse knowledge. Understanding interaction information requires building a model of the current state of the conversation (e.g. the current speaker and addressee). Generating propositional information requires a planning process presenting multi-sentence output and managing the order of presentation of interdependent facts.

4. *Conversational Function Model*: Explicit represention of conversational functions provides both modularity and a principled way to combine different modalities. The core process of the system might operate independently on functions rather than surface level representation, such as sentences, while other modules at input and output of the system translate input into functions, and functions into outputs.

They designed the REA (Real Estate Agent) architecture fulfilling the requirement above. REA is a computer generated humanoid with graphical body, which can deal with speech input and output, facial display and gestural output.

The architecture includes: *Input Manager*, *Understanding Module*, *Reaction Module*, *Response Planner Module*, *Generation Module*, *Action Scheduling Module*, and *Hardwired Reaction*. The *Input Manager* gets multimodal inputs and decides whether the data requires instant reaction or deliberate discourse processing. The *Understanding Module* interprets all input modalities into a abstracted understanding of what the user is doing. It receives inputs from the *Input Manager* and accesses knowledge about the application domain and the current discourse context. The *Reaction Module* select the action to perform, which receives asynchronous updates from the *Input Manager* and *Understanding Module*, and uses information about the domain and current discourse state to determine the action to select. The *Response Planner* formulates sequences of actions to achieve desired communicative or task goals. The *Generation Module* realizes discourse functions output from the *Reaction Module* by producing a set of coordinated primitive actions, sending the actions to the *Action Scheduler* for performance, and monitoring their execution. The *Action Scheduler* schedules motor events to be sent to the animation rendering engine. A crucial function of the scheduler is to prevent collisions between competing motor requests. The *Hardwired Reaction* handles spontaneous reaction to stimuli (e.g. eye gaze). Table 2.1 shows Rea's interactional output behaviors.

2.3 Framework for Multiparty Conversation

Many researchers presented differences between two-participant interactions and multiparty interactions, and dyadic models cannot be easily applied to multiparty cases. Major differences can be observed as ways of *ratification*, *addressing* and *engagement*, which compose some elements of the "participation structure."

2.3.1 Participation Role and Ratification

Goffman defined distinctively "ratified" and "unratified" participants in a conversation. Ratified participants are participants who "have declared themselves officially open to one another for purposes of spoken

Table 2.1: Rea's interactional output behaviors.

State	Output Function	Behaviors
User Present	Open interaction	Look at user. Smile. Toss head.
	Attend	Face user.
	End of interaction	Turn away
	Greet	Wave. Say "hello".
Rea Speaking	Give turn	Relax hands. Look at user. Raise eyebrows
	Signoff	Wave. Say "bye"
User Speaking	Give feedback	Nod head, paraverbal ("hmm")
	Want turn.	Look at user. Raise hands.
	Take turn.	Look at user. Raise hands to begin

communication and guarantee together to maintain a flow of word" (Goffman, 1967). And unratified participants are treated as a person who are not not formally participating in the conversation. The ratification is a collaborative process conducted by participants in the conversation. Goffman classified three kinds of hearers: *addressed recipients*, *unaddressed recipients* and *bystanders*. While addressed and unaddressed recipients are ratified participants in the conversation, unaddressed recipients are participants who are not being specifically addressed by the speaker. Bystanders are unratified participants in the conversation but is perceivable by the ratified participants. Drawing Goffmann's categorization, Clark proposed a modified participation structure. Participants in the action are *speaker*, *addressees*, and *side-participant* who is participating the conversation but currently not being addressed. And all other listeners who are not in the conversations are called *overhearers*. In the Clark's categorization, overheares are not ratified by the speaker. Overheares can be divided into *bystanders* and *eavesdroppers*. Bystanders are participants who are present a same place with the speaker's awareness but they are not part of the conversation. Eavesdroppers are participants who are listening to conversation without the speaker's awareness.

2.3.2 Addressing and Recipient Design

In contrast to dyadic, each action of a speaker in a group assigns other participants' roles by its *recipient design*. The *role assignment* is a critical part of multiparty conversation framework. Sacks et al. described the recipient design as "a multitude of respects in which the talk by a party in a conversation is constructed or designed in ways which display an orientation and sensitivity to the particular other(s) who are co-participants" (Sacks et al., 1974). Levinson also investigated the recipient design specifically for indirectly targeted utterances focusing on targeted participants who are not being addressed (Levinson, 1988). Clark et al. presented a concept of *audience design* to extend the concept of recipient design to include overhearers in addition to other ratified participants (Clark and Carlson, 1982). A speaker designs its utterances in different ways based on target hearers. He divided the concept of audience design into participants design, addressee design and overhearers design. In participant design, the speaker intends participants are informed about an utterance that the speaker is performing towards addressees and to recognize the meaning of the action. During designing an utterance, the speaker primarily pays an attention on addressees (addressee design), and he/she does not regard overhearers understand an utterance (overhearers design).

2.3.3 Engagement

Sidner et al. dealt with engagement in multimodal ways, including eye gaze. They defined engagement as "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake" (Sidner et al., 2004). This process includes: (1) initial contact, (2) negotiating a collaboration, (3) checking that other is still taking part in the interaction, (4) evaluating whether to stay involved, and (5) deciding when to end the connection. Martin et al. proposed the appraisal theory that is concerned with the interpersonal in language, with the subjective presence of writers/speakers in texts as they adopt stances towards both the material they present and those with whom they communicate. It encompasses three sub-categories, namely *Attitude*, *Engagement*, and *Graduation* (Martin and White, 2005). *Attitude* deals with expressions of affect, judgement, and appreciation. *Engagement* focuses on language use by which speakers negotiate an interpersonal space for their positions and the strategies which they uses to either acknowledge, ignore, or curtail other voices or points of view. *Graduation* focuses on the resources by which sparkers regulate the impact of these resources.

In terms of the way of controlling engagement, Whittaker et al. analyzed two-participant dialogues to investigate the mechanism how each control was signaled by speakers and how it affects discourse structure, including the lower control level, topic level and global organization level (Whittaker and Stenton, 1988). They found that utterance type predicted shifts of control. control is a useful parameter for identifying discourse structure. Using this parameter they identified three levels of structure in the dialogues: (a) control phases, (b) topic and (c) global organization. For the control level, they found that three types of utterances (prompts, repetitions and summaries) were consistently used to signal for controlling. For the topic level, they found that interruptions introduce a new topic. And the global organization is organized by topic initiation and controls.

2.4 Layered Model of Conversational Processes and Protocols

In order to capture and model multiparty dynamics, with taking components above into account, layered models of conversational processes have been proposed. Drawing on Clark's model of language use Bohus et al. presented open-world dialog (Clark and Schaefer, 1989; Clark, 1996). In Clark's model, natural language interaction is regarded as a joint activity where participants in a conversation attend to each other and coordinate their actions on several different levels to establish and maintain mutual ground. Components the Clark's concept includes *channel level*, *signal level*, *intention level* and *conversation level*. At the lowest level (*channel*), the participants might coordinate their actions to establish, maintain or break an open communication channel. At the second (*signal*) level, participants might coordinate the presentation and recognition of various communicative signals. At the third (*intention*) level, participants might coordinate to correctly interpret the meaning of these signals. Finally, at the fourth (*conversation*) level, participants might coordinate and plan their overall collaborative activities and interaction. Bohus et al. closely studied the *channel level* for managing engagement (Bohus and Horvitz, 2009a,b) and *signal level* for turn-taking Bohus and Horvitz (2011b). The engagement management prototype has situational awareness capabilities of face detection and tracking (a multiple face detector and tracker for detecting and tracking the location of each user), pose tracking (pose tracker provides 3D head orientation information for each engaged user), focus of attention (a direct conditional model was used to infer whether the attention of each user in the scene is oriented towards the system or not), agent characterization (a simple conditional model of users), group inferences (a pairwise analysis of the agents in the scene to infer group relationships).

Kobayashi et al. described conversational protocol model as an analogy of a network model (the Open

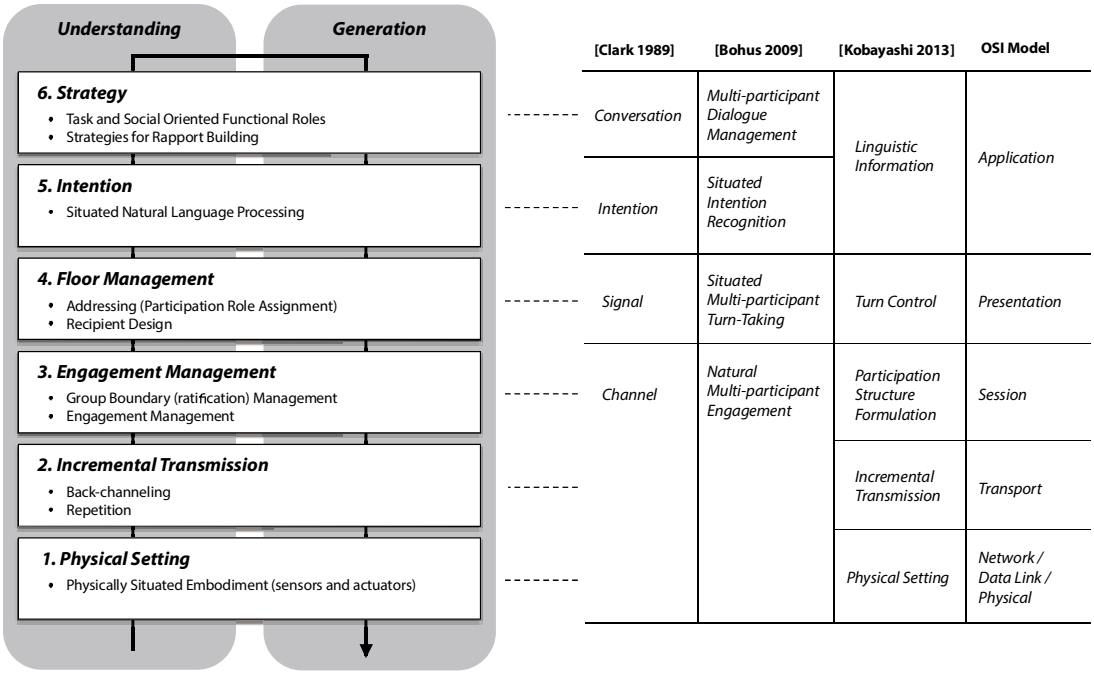


Figure 2-1: Layered model of conversational processing and protocols. The left is our model, corresponding to models of Clark, Bohus et al. Kobayashi et al., as well as the OSI model in the right.

Systems Interconnection (OSI) model[1]) in order to realize a "frustration-free" communication, with taking paralinguistic signals into account (Kobayashi and Fujie, 2013; Kobayashi et al., 2014). They defined protocols as "sets of rules for achieving reliable and efficient information transmission." The OSI model is a conceptual model that characterizes and standardizes the internal functions of a communication system by partitioning it into seven abstraction layers: *physical layer*, *data link layer*, *network layer*, *transport layer*, *session layer*, *presentation layer* and *application layer*[2] . Based on the abstraction of the OSI model, they categorized the protocols into four layers: *physical layer*, *message-transmission layer*, *turn-taking layer*, *group conversation layer*. The *message-transmission layer* is responsible for incremental communication. For example, acknowledgment, maintaining a current turn without a turn shift, can harmonize the rhythm of the conversation by helping the speaker to speak with appropriate timing (Ward and Tsukahara, 2003; Kitaoka et al., 2005). They considered use of paralinguistic information for recognizing acknowledgments using prosody information of an user's back-channel feedback and repetition (Fujie et al., 2006), as well as generating acknowledgments with a content plan using a finite state machine (FSM) representation (Fujie

[1]https://www.iso.org/obp/ui/#iso:std:iso-iec:7498:-1:ed-1:v2:en

[2]*Physical layer* is the first layer that defines physical connections including electrical, mechanical, functional and procedural specifications. *Data link layer* is the second layer that provides a reliable data transfer between directly connected nodes. *Network layer* is the third later that has responsibilities routing variable length data sequences from end to end. *Transport layer* is the forth layer that has responsibilities transferring variable length data sequences from a source to a destination host via one or more networks with error recovery and retransmission capabilities (this layer uses the Transmission Control Protocol (TCP) that built on top of the Internet Protocol (IP)). *Session layer* is the fifth layer that establishes, manages and terminates connections between computer nodes. *Presentation layer* is the sixth layer that has responsibilities to deliver and format information to facilitate differences in data representation for the application layer. *Application layer* is the seventh layer that is the closest to the end user.

et al., 2004). For the *turn-taking layer*, based on their findings that differences between keeping and releasing turns appear in the linguistic representation, prosody, and eye-gaze near the final part of the utterance or phrase boundary[3], they proposed an integrated model of turn-taking using paralinguistic information. They also used user's internal states: "expectation"(how an user/system is expected to take a turn) and "willingness"(how an user/system is willing to take a turn). While those layers take care of dyadic interaction, the *group conversation layer*, takes care of participation structure in a multiparty situation. They presented a turn-taking model in three-participant situation, including a robot, using facial direction and voice activity information (Matsusaka et al., 2003)

Drawing on the these attempts to describe conversational model in terms of protocols, we present a layered model of conversational processing and protocols, as shown in Figure 2-1.

2.4.1 Facilitation Strategies

Zhao et al. proposed a computational model of rapport management in a dyadic conversation. They reviewed existing literature and their corpus of peer tutoring data to develop a framework able to explain how humans in dyadic interactions build, maintain, and break rapport through the use of specific conversational strategies that function to fulfill specific social goals, and that are instantiated in particular verbal and nonverbal behaviors (Zhao et al., 2014). They also proposed an architecture to realize the rapport model (Papangelis et al., 2014). However, strategical decision making in dyadic and multiparty conversations are quite different phenomena. Much research have been conducted in social psychology and sociology on small group dynamics. In this research, interactions in small groups have been described as *role* (Benne and Sheats, 1948; Biddle, 1979, 1986; Salazar, 1996; Gatica-Perez, 2009; Hare, 1994; Pianesi et al., 2008; Zancanaro et al., 2006; Dong and Pentland, 2007; Dong et al., 2007, 2013).

Functional Roles in Small Groups

Benne et al. analyzed functional roles in small groups to understand the activities of individuals in small groups (Benne and Sheats, 1948). They categorized functional roles in small groups into three classes: (1) *Group task roles*, (2) *Group building and maintenance roles*, and (3) *Individual roles*. The *Group task roles* are defined as "related to the task which the group is deciding to undertake or has undertaken," whose roles address concerns about the facilitation and coordination activities for task accomplishment. The *Group building and maintenance roles* are defined as "oriented toward the functioning of the group as a group," which contribute to social structures and interpersonal relations. Finally, the *Individual roles* are directed toward the individual satisfaction of each participant's individual needs. They deal with individual goals that are not relevant either to the group task or to group maintenance. Table 2.2, 2.3, 2.4 show Benne's categorization of group task roles, group building and maintenance roles, and individual roles, respectively.

Drawing on Benne's work, Bales proposed interaction process analysis (IPA), a framework for the classification of individual behaviors in group interaction in a two-dimensional role space consisting of a *Task area* and a *Socio-emotional area* (Bales, 1950). The Bale's IPA scheme is along twelve interaction categories as is shown in Table 2.5. Six of these interaction categories correspond to instrumental (task-related) interaction and the 38 other six correspond to expressive (social-emotional) interaction. The roles related to the *Task area* concern behavioral manifestations that impact the management and solution of problems that a group is addressing. Examples of task-oriented activities include initiating the floor, giving

[3]To express intent to keep the turn, the speaker usually maintains a higher voice pitch and does not gaze at a particular person. These behaviors create an atmosphere in which a hearer hesitates to speak. To express willingness to release (hand over) the turn, the speaker usually lowers his or her voice pitch and gazes at a particular hearer.

Table 2.2: Benne's Categolization of group task roles.

Category	Description and Patterns
Initiator-Contributor	"suggests or proposes to the group new ideas or a changed way of regarding the group problem or goal. The novelty proposed may take the form of suggestions of a new group goal or a new definition of the problem. It may take the form of a suggested solution or some way of handling a difficulty that the group has encountered. Or it may take the form of a proposed new procedure for the group, a new way of organizing the group for the task ahead."
Information Seeker	"asks for clarification of suggestions made in terms of their factual adequacy for authoritative information and acts pertinent to the problem being discussed."
Initiator-Contributor	"suggests or proposes to the group new ideas or a changed way of regarding the group problem or goal. The novelty proposed may take the form of suggestions of a new group goal or a new definition of the problem. It may take the form of a suggested solution or some way of handling a difficulty that the group has encountered. Or it may take the form of a proposed new procedure for the group, a new way of organizing the group for the task ahead."
Information Seeker	"asks for clarification of suggestions made in terms of their factual adequacy for authoritative information and acts pertinent to the problem being discussed."
Opinion Seeker	"asks not primarily for the facts of the case but for a clarification of the values pertinent to what the group is undertaking or of values involved in a suggestion made or in alternative suggestions."
Information Giver	"offers facts or generalizations that are authoritative or relates his own experience pertinently to the group problem."
Opinion Giver	"states his belief or opinion pertinently to a suggestion made or to alternative suggestions. The emphasis is on his proposal of what should become the group's view of pertinent values, not primarily upon relevant facts or information."
Elaborator	"spells out suggestions in terms of examples or developed meanings, offers a rationale for suggestions previously made and tries to deduce how an idea or suggestion would work out if adopted by the group."
Coordinator	"shows or clarifies the relationships among various ideas and suggestions, tries to pull ideas and suggestions together or tries to coordinate the activities of various members or sub-groups."
Orienter	"defines the position of the group with respect to its goals by summarizing what has occurred, points to departures from agreed upon directions or goals, or raises questions about the direction which the group discussion is taking."
Evaluator Critic	"subjects the accomplishment of the group to some standard or set of standards of group-functioning in the context of the group task. Thus, he may evaluate or question the practicality, the logic, the facts or the procedure of a suggestion or of some unit of group discussion."
Energizer	"prods the group to action or decision, attempts to stimulate or arouse the group to greater or higher quality activity."
Procedural Technician	"expedites group movement by doing things for the group-performing routine tasks, e.g., distributing materials, or manipulating objects for the group, e.g., rearranging the seating or running the recording machine, etc."
Recorder	"writes down suggestions, makes a record of group decisions, or writes down the product of discussion. The recorder role is the group memory."

information, and providing suggestions regarding a task. The roles related to the *Socio-emotional area* affect the interpersonal relationships either by supporting, enforcing, or weakening them. For instance, complementing another person to increase group cohesion and mutual trust among members is one example of positive socio-emotional behavior.

Patterns of Facilitation Strategies and Skills

In order to capture facilitation skills, there are some attempts to use *pattern languages*. A *pattern language*, originally proposed by Christopher Alexander (Alexander et al., 1977), is a method of universally describing good design practices so that ordinary people can use it to solve large and complex design problems. As for facilitation skills of group process, The Dialogue and Deliberation, Group Pattern Language Project (NCDD) presented pattern languages of group process as the results of open discussions[4]. Kahn et al. reported pattern languages for social robots (Kahn et al., 2008). Shimizu et. al proposed 56 facilitation patterns for designing an experiential learning program. They applied pattern language to the design and

[4]http://groupworksdeck.org/

Table 2.3: Benne's Categolization of group building and maintenance roles.

Category	Description and Patterns
Encourager	"praises, agrees with and accepts the contribution of others. He indicates warmth and solidarity in his attitude toward other group members, offers commendation and praise and in various ways indicates understanding and acceptance of other points of view, ideas and suggestions."
Harmonizer	"mediates the differences between other members, attempts to reconcile disagreements, relieves tension in conflict situations through jesting or pouring oil on the troubled waters, etc."
Compromiser	"operates from within a conflict in which his idea or position is involved. He may offer compromise by yielding status, admitting his error, by disciplining himself to maintain group harmony, or by coming half-way in moving along with the group."
Gate-Keeper and Expediter	"attempts to keep communication channels open by encouraging or facilitating the participation of others ("*we haven't got the ideas of Mr. X yet*," etc.) or by proposing regulation of the flow of communication ("*why don't we limit the length of our contributions so that everyone will have a chance to contribute?*", etc.)"
Standard Setter / Ego Ideal	"expresses standards for the group to attempt to achieve in its functioning or applies standards in evaluating the quality of group processes."
Group-Observer and Commentator	"keeps records of various aspects of group process and feeds such data with proposed interpretations into the group's evaluation of its own procedures."
Follower	"goes along with the movement of the group, more or less passively accepting the ideas of others, serving as an audience in group discussion and decision."

Table 2.4: Benne's Categolization of individual roles.

Category	Description and Patterns
Encourager	"praises, agrees with and accepts the contribution of others. He indicates warmth and solidarity in his attitude toward other group members, offers commendation and praise and in various ways indicates understanding and acceptance of other points of view, ideas and suggestions."
Harmonizer	"mediates the differences between other members, attempts to reconcile disagreements, relieves tension in conflict situations through jesting or pouring oil on the troubled waters, etc."
Compromiser	"operates from within a conflict in which his idea or position is involved. He may offer compromise by yielding status, admitting his error, by disciplining himself to maintain group harmony, or by coming half-way in moving along with the group."
Gate-Keeper and Expediter	"attempts to keep communication channels open by encouraging or facilitating the participation of others ("*we haven't got the ideas of Mr. X yet*," etc.) or by proposing regulation of the flow of communication ("*why don't we limit the length of our contributions so that everyone will have a chance to contribute?*", etc.)"
Standard Setter / Ego Ideal	"expresses standards for the group to attempt to achieve in its functioning or applies standards in evaluating the quality of group processes."
Group-Observer and Commentator	"keeps records of various aspects of group process and feeds such data with proposed interpretations into the group's evaluation of its own procedures."
Follower	"goes along with the movement of the group, more or less passively accepting the ideas of others, serving as an audience in group discussion and decision."

facilitation of experiential learning program. They reported that facilitators could choose the solution from the facilitation pattern in a coordinated fashion using the facilitation patterns (Shimizu and Iba, 2006).

2.5 Computational Architecture for Facilitation Robots

2.5.1 Cognitive Architecture: Declarative and Procedural Memories

Newell argues for the need of a set of general assumptions for cognitive models that account for all of cognition: a unified theory of cognition (UTC) (Newell, 1994). Anderson defined cognitive architecture as "a specification of the structure of the brain at a level of abstraction that explains how it achieves the function of mind." (Anderson, 2007). Inspired Newell's work, Anderson proposed and implemented ACT-R (Adaptive Control of Thought - Rational) (Anderson et al., 2004; Anderson, 2007). The most important

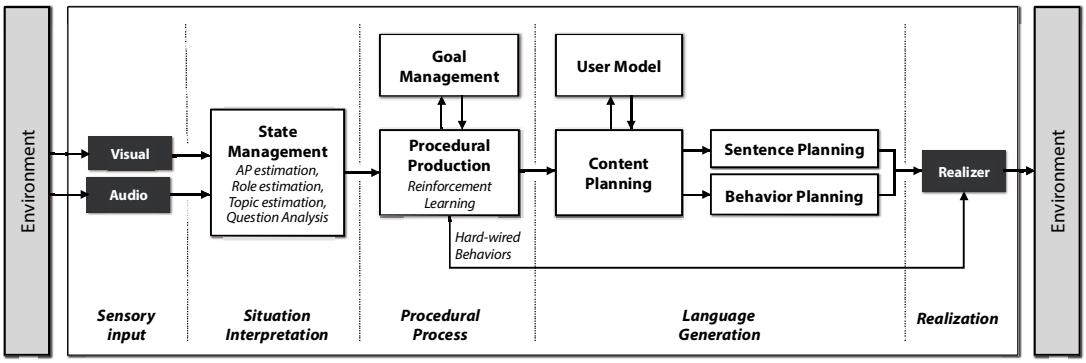
Table 2.5: Bales' Interaction Process Analysis.

Category	Behaviors
Positive Expressive Interaction Categories (*Socio-emotional area*)	*Shows Solidarity*
	Shows Tension Release
	Agrees
Instrumental Interaction Categories (*Task area*)	*Gives Suggestion*
	Gives Opinion
	Gives Orientation
	Asks for Orientation
	Asks for Opinion
	Asks for Suggestion
Negative Expressive Interaction Categories (*Socio-emotional area*)	*Disagrees*
	Shows Tension
	Shows Antagonism

assumption of ACT-R is that human knowledge can be divided into *declarative* (fact-based) and and *procedural* (rule-based) memories. *Declarative* memory is a type of memory consisting of facts. *Procedural* memory is a type of long-term memories about how we do things, including motor skills. ACT-R consists of a set of modules, each module processes a different kind of information. A visual sensory module identifies objects in the field, a manual motor module controls the hands, a declarative module retrieves information from declarative memory, and a goal module keeps track of current task goals and intentions. A procedural system coordinates these modules and produces behaviors. Production rules proceeded in the central production system represent ACT-R's procedural memory. Trafton et al. proposed integrating vision and audition within a cognitive architecture to track conversations (Trafton et al., 2008, 2009, 2013).

2.5.2 SCHEMA Framework: Architecture for Facilitation Robots

Based on the requirements and elements of facilitation model, as well as the general concepts of cognitive architectures we reviewed above, we propose a computational architecture for multiparty conversation facilitation robots, namely the SCHEMA Framework. The SCHEMA Framework mainly consists of the following processes: the *Perception Process* the *Procedural Production Process* the *Language Generation Process*. The *Perception Process* process interprets situations based on visual and auditory information. This process includes Adjacency Recognition, Participation Recognition, Topic Recognition and Question Analysis. Each time the system detects an endpoint of participant's speech from the automatic speech recognition (ASR) module, it interprets the current situation. This process will be described in Chapter 3 in detail. The *Procedural Production Process* produces procedural actions to manage a group, referred Goal Management Module. This module is modeled as a reinforcement learning framework (partially observable Markov decision process (POMDP)). This process will be described in Chapter 3 in detail. The *Language Generation Process*. is divided into factoid and non-factoid typed answer generation modules. The factoid typed answer generation module refers to structured knowledge databases organized using Semantic Web techniques. The non-factoid typed answer generation module generates the system's own opinions automatically extracted from a large indefinite number of reviews on the Web. It also has an utterance combination mechanism that combines factoid and non-factoid typed responses to realize the additional phrasing function. This process will be described in Chapter 4 in detail.



Situation Interpretation
interprets situations based on visual and auditory information. This process includes Adjacency Recognition, Participation Recognition, Topic Recognition and Question Analysis. The details will be described in Chapter 3.

Procedural Process *(Group Maintenance)*
produces procedural actions to manage a group, referring Goal Management Module. This module is modeled as a reinforcement learning framework (partially observable Markov decision process (POMDP)). The details will be described in Chapter 3.

Language Generation *(SCHEMA QA)*
has Content Planning, Microplanning (Sentence and Behavior Planning) and Realization processes. The details will be described in Chapter 4.

Figure 2-2: Whole architecture of the SCHEMA Framework.

2.6 Conclusions

In this section, we described the SCHEMA Framework, a general architecture for multiparty facilitation robots. Drawing the related works in the fields of cognitive architectures and situated conversational agents, as well as group process analysis, we described requirements for facilitation robots, including (1) multimodal situation awareness, (2) strategic decision making and (3) semantic understanding and generation. Then, we described a whole of architecture for facilitation process.

In the following sections, we will discuss each module in details. In Section 3, we will present present facilitation strategies for group maintenance that fulfills the first and second requirements. In Section 4, we will present sentence understanding and generation process that fulfills the third requirement. In Section 5, we will describe an implementation of the SCHEMA Framework, and finally, in Section 6, we will propose an application using the SCHEMA framework.

"The greatest happiness of the greatest number is the foundation of morals and legislation."

Jeremy Bentham

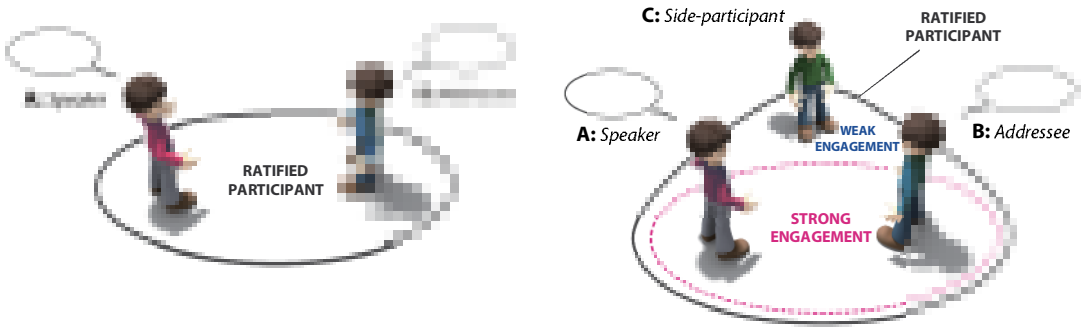
3 Engagement Density Control

In this chapter, we present a framework for facilitation robots that regulate imbalanced engagement density in four-participant conversation as the fourth participant with proper procedures for obtaining initiatives. Four is the special number in multiparty conversations. In three-participant conversations, the minimum unit for multiparty conversations, social imbalance, in which a participant is left behind in the current conversation, sometimes occurs. In such scenarios, a conversational robot has the potential to objectively observe and control situations as the fourth participant. Consequently, we present model procedures for obtaining conversational initiatives in incremental steps to harmonize such four-participant conversations. During the procedures, a facilitator must be aware of both the presence of dominant participants leading the current conversation and the status of any participant that is left behind. We model and optimize these situations and procedures as a partially observable Markov decision process (POMDP), which is suitable for real-world sequential decision processes. The results of experiments conducted to evaluate the proposed procedures show evidence of their acceptability and feeling of groupness.

3.1 Introduction

We present a framework for facilitation robots that regulates imbalanced engagement density in four-participant conversation as the fourth participant with proper procedures for obtaining initiatives. Four is the special number in multiparty conversations. The three-participant conversation is the minimum unit where the participants autonomously organize a multiparty conversational situation. The fourth participant is the first person who can objectively observe the conversational situation. In three-participant conversations, social imbalance, in which a participant is left behind in the current conversation, sometimes occurs. In such scenarios, a conversational robot has the potential to objectively observe and control situations as the fourth participant. A four-participant conversational situation, where three participants and a facilitator are participating, is the minimum unit of the facilitation process model.

Figure 3-1 (a) depicts a two-participant conversation. In such situations, conversational context, includ-



(a) 2-participant (b) 3-participant

Figure 3-1: (a) Two-participant conversation model, which have been focused upon by conventional dialogue systems. (b) Three-participant conversation model; the minimum unit for a multiparty conversation.

ing engagement (Sidner et al., 2004) and turn-taking (Sacks et al., 1992), is commonly grounded between two interlocutors. Many dialogue systems have dealt with turn-taking within two-participant engagement (Raux and Eskenazi, 2009; Chao and Thomaz, 2012b). However, in three-participant conversations as shown in Figure 3-1 (b), which is the minimum unit for multiparty conversation, engagement and turn-taking cannot always be identified among the participants. In terms of turn-taking in multiparty conversations, the participation structure model was presented by Clark (Clark, 1996), drawing on Goffman's work (Goffman, 1981). In the participation structure model, each participant is assigned a participation role considered by the current speaker, where *speaker*, *addressee*, and *side-participant* are "ratified participants" (Goffman, 1981). In such three-participant situations, interactions between two dominant participants primarily occur between participants A and B, and the other participants, who cannot properly get the floor to speak for a long while (can neither be promoted to a speaker nor an addressee), tends to get left behind, even though all participants are ratified.

Such a social imbalance problem cannot be solved easily because participation roles do not always share common ground among the ratified participants. For example, in Figure 3-1 (b), participant C might not be able to properly take chances to assume the floor to speak for a while, and thus, from his viewpoint, is left out of the dominant conversation, even though floor exchanges may be well maintained among participants from participant A's viewpoint. If situational comprehension of the participation structure is diverged among the participants and participant A cannot recognize the left-behind situation, he may not be motivated to self-initiate control of the situation. In the left-behind situation, the engagement density may be different between dominant participants and the left-behind participant. The dominant participants' engagement is so strong that participant C's engagement with others is relatively weak. In addition, it is also possible that participant C cannot share a common interest topic with the other participants. Consequently, socially imbalanced three-participant situations dictate the need for an additional *facilitator* participant to help the left-behind participant "harmonize" with the other participants. In this context, "harmonize" means maintaining equality of engagement density within the group. A four-participant conversational situation is the minimum unit of the facilitation process model, which has never been discussed substantially in research of both conversational analysis and dialogue systems. Conversational robots have the potential to participate as the fourth participant to facilitate such conversations, as is illustrated in Figure 3-2. Kobayashi et al. have discussed the importance of situated human-like conversational robots, which are capable of omitting and understanding conversational protocols (Kobayashi and Fujie, 2013). Generally, when a facilitator (robot)

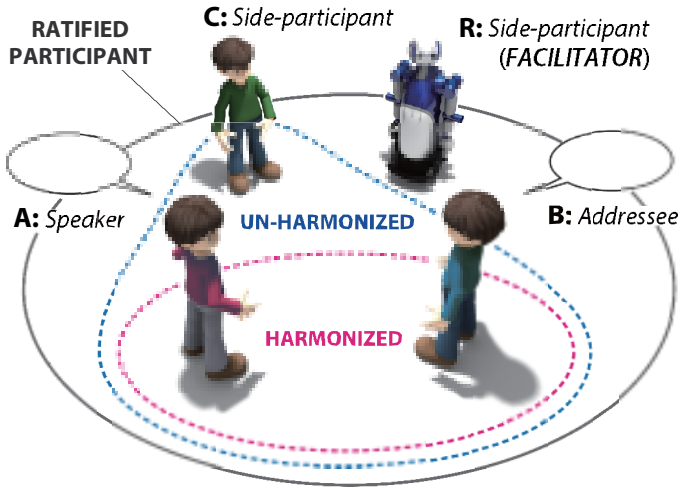


Figure 3-2: Four-participant conversation; the minimum unit of conversation that needs facilitation process. A facilitator (robot) can objectively observe the situation, and regulate imbalanced situations with proper procedural steps. In this case, person C is left behind so the robot is trying to approach him with being aware of the presence of dominant participants (A and B) leading the current conversation.

steps into the situation to coordinate, it should follow properly established procedures to obtain initiative within situations and give this initiate back to the other participants. To coordinate situations, a facilitator must take the following procedural steps. (1) Be aware of both the presence of dominant participants leading the current conversation and the status of a left-behind participant; (2) obtain an initiative to control the situation and wait for approval from the others, either explicitly or implicitly; and (3) give the floor to a suitable participant (sometimes by initiating a new topic).

Various related research on specially situated facilitation agents in multiparty conversations has been conducted. Matsusaka et al. pioneered the use of a physical robot participating in multiparty conversations (Matsusaka et al., 2003). We have previously developed a multiparty quiz-game-type facilitation system for elderly care (Matsuyama et al., 2008) and reported the effectiveness of the existence of a robot (Matsuyama et al., 2010). Dosaka et al. developed a thought-evoking dialogue system for multiparty conversations with a quiz-game task (Dohsaka et al., 2009). They reported that the existence of agents and empathic expressions is effective for user satisfaction and can increase the number of user utterances. Sidner et al. developed an agent system that can engage with users, where they defined engagement as "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake" (Sidner et al., 2004). Bohus et al. modeled engagement in multiparty conversations using Sinder's definition, i.e., open world dialogue (Bohus and Horvitz, 2009a). They evaluated the effectiveness of multimodalities, including gaze, gesture, and speech, for a multiparty conversation facilitating agent (Bohus and Horvitz, 2010b). In terms of facilitation, Kumar et al. designed a dialogue action selection model based on Bales' Socio-Emotional Interaction Categories for text-based character agents (Kumar et al., 2011). However, there is a lack of profound consideration regarding engagement density in multiparty conversational situations and procedural operations for obtaining initiative to control conversational situations while considering their side-effects, which typically occur in multiparty conversational situations.

In this paper, we propose a procedural facilitation process framework to harmonize a four-participant conversational situation. The situations and procedures are modeled and optimized as a partially observable Markov decision process (POMDP), which is suitable for real-world sequential decision processes, including dialogue systems (Williams and Young, 2007). The remainder of this paper is organized as follows. We begin by reviewing facilitation frameworks in small groups and describing procedures for maintaining small groups. In Section 3.3, we discuss how to model them as POMDP. In Section 3.4, we give an overview of the architecture of our proposed system. We then discuss three experiments conducted to verify the efficacy of the small group maintenance procedures and the performance of POMDP. Finally, we summarize and conclude this study.

3.2 Theoretical Framework for Engagement Control

In this section, in order to organize the facilitation framework, at first, we review related works of facilitation models in small groups, specifically functional roles of group members that have been defined to analyze facilitation processes. Then we review engagement models, and we propose the harmony model.

3.2.1 Small Group Maintenance

Benne et al. analyzed functional roles in small groups to understand the activities of individuals in small groups (Benne and Sheats, 1948). They categorized functional roles in small groups into three classes: *Group task roles*, *Group building and maintenance roles*, and *Individual roles*. Table 3.1 shows the Benne's categorization of functional roles. The *Group task roles* are defined as "related to the task which the group is deciding to undertake or has undertaken," whose roles address concerns about the facilitation and coordination activities for task accomplishment. The *Group building and maintenance roles* are defined as "oriented toward the functioning of the group as a group," which contribute to social structures and interpersonal relations. Finally, the *Individual roles* are directed toward the individual satisfaction of each participant's individual needs. They deal with individual goals that are not relevant either to the group task or to group maintenance. Drawing on Benne's work, Bales proposed interaction process analysis (IPA), a framework for the classification of individual behavior in a two-dimensional role space consisting of a *Task area* and a *Socio-emotional area* (Bales, 1950). The roles related to the *Task area* concern behavioral manifestations that impact the management and solution of problems that a group is addressing. Examples of task-oriented activities include initiating the floor, giving information, and providing suggestions regarding a task. The roles related to the *Socio-emotional area* affect the interpersonal relationships either by supporting, enforcing, or weakening them. For instance, complementing another person to increase group cohesion and mutual trust among members is one example of positive socio-emotional behavior.

In this research, we employ Benne's *Group building and maintenance roles*, which are related to Bales's *Socio-emotional area*, in order to arrange the following three abstract functional roles of group maintenance:

1. *Observation Role*: Overlooking the conversation situation by finding appropriate topics, observing the motivations and moods of the participants, and comprehending the relations between participants in conversations. This person follows the conversation and comments and interprets the group's internal process. This role inherits *Observer and commentator* and *Encourager*.

2. *Floor Maintenance Role*: Maintaining the chance for the floor in the group in a direct/indirect way. This person encourages or asks questions of the person who is not or could not get engaged in con-

Table 3.1: Benne's categorization of functional roles (Benne and Sheats, 1948).

Category	Functional roles
Group task roles	*Initiator-contributor*, *Information seeker*, *Opinion seeker*, *Information giver*, *Elevator*, *Coordinator*, *Orienter*, *Elevator-critic*, *Energizer* *Procedural technician*, *Recorder*
Group building and maintenance roles	*Compromiser*, *Harmonizer*, *Standard setter*, *Gatekeeper and expediter*, *Encourager*, *Observer and commentator* *Follower.*
Individual roles	*Aggressor*, *Blocker*, *Recognition-seeker*, *Self-confessor*, *Playboy* , *Dominator*, *Help-seeker*, *Special interest pleader*

versations, and attempts to keep the communication channel open. This role inherits *Gatekeeper*, *Expediter*, and *Encourager*.

3. *Topic Maintenance Role*: Maintaining for conflict, ideas, and topics. This person mediates the difference between other members, attempts to reconcile disagreements, and relieves tension in conflict situations. This role inherits *Compromiser*, *Harmonizer*, and *Standard setter*.

As we described in Section 4.1, a facilitator must take the following procedural steps:

1. (**Observation**) Be aware of both the presence of dominant participants leading the current conversation and the status of a participant who is left behind (*Observation Role*)

2. (**Obtaining an initiative**) Obtain an initiative to control the situation and wait for approval from the others, either explicitly or implicitly

3. (**Floor and topic maintenance**) Give a floor to a suitable participant, sometimes with initiating a new topic (*Floor Maintenance Role* and *Topic Maintenance Role*)

Here, the *Observation Role* represents the first step, and *Floor Maintenance Role* and *Topic Maintenance Role* represent the third process. Before invoking the *Floor Maintenance Role* and *Topic Maintenance Role*, a facilitator obtains an initiative in the second step. Such a procedure obtaining an initiative has never been substantially discussed in past works including the Benne and Bales's literature. We will formalize the procedure in more depth below.

3.2.2 Engagement Density

In order to formalize procedural steps obtaining an initiative controlling a situation, we begin by extending the participation structure model in multiparty conversations. The participation structure model was presented by Clark (Clark, 1996), drawing on Goffman's work (Goffman, 1981). In this model, each participant is assigned a participation role considered by the current speaker, where *speaker*, *addressee*, and *side-participant* are "ratified participants." Ratified participants include the speaker and addressees, as well as a side-participant who is taking part in the conversation but is not currently being addressed. All other listeners, who we refer to as over-hearers, have no rights or responsibilities within the structure. *Over-hearers* come in two main types. *Bystanders* are those who are openly present but not part of the conversation.

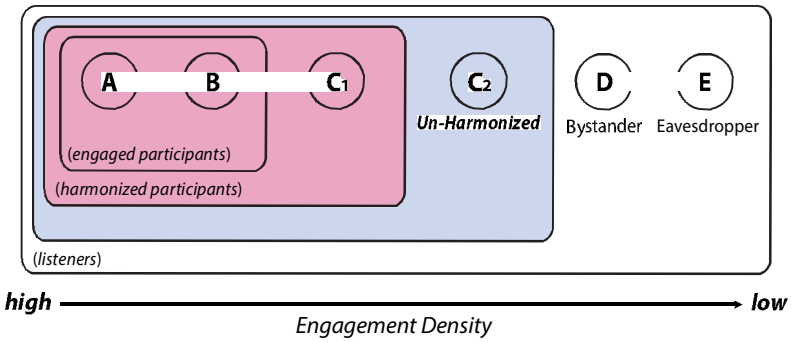


Figure 3-3: Participation structure model extended from Clark's model. Speaker, Addressee and Side-participant are "ratified participants". Not ratified participants are divided into two types: Bystanders and Eavesdroppers. Side-participants are divided into two types: *harmonized* side-participant and *un-harmonized* side-participant according to their engagement density.

Eavesdroppers are those who listen in without the speaker's awareness. The *speaker* must pay close attention to these distinctions when speaking. For example, the *speaker* must distinguish *addressee* from *side-participants*. When the *speaker* asks an *addressee* a question, the *speaker* must make sure that it is the *addressee* who is intended to answer the question, and not *side-participants*. However, the *speaker* must also ensure that the *side-participant* understands the question directed at the *addressee*. In addition, the *speaker* must consider the *over-hearers*. However, because the *over-hearers* have no rights or responsibilities in the current conversation, the *speaker* can treat them as he pleases.

In this paper, we extend Clark's model with the concept of engagement. In terms of engagement among conversational participants, Martin et al. proposed the appraisal theory that is concerned with the interpersonal in language, with the subjective presence of writers/speakers in texts as they adopt stances towards both the material they present and those with whom they communicate. It encompasses three sub-categories, namely *Attitude*, *Engagement*, and *Graduation* (Martin and White, 2005). *Attitude* deals with expressions of affect, judgement, and appreciation. *Engagement* focuses on language use by which speakers negotiate an interpersonal space for their positions and the strategies which they uses to either acknowledge, ignore, or curtail other voices or points of view. *Graduation* focuses on the resources by which sparkers regulate the impact of these resources. Sidner et al. dealt with engagement in multimodal ways, including eye gaze. They defined engagement as "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake" (Sidner et al., 2004). This process includes: (1) initial contact, (2) negotiating a collaboration, (3) checking that other is still taking part in the interaction, (4) evaluating whether to stay involved, and (5) deciding when to end the connection. Based on these previous studies, we define engagement as the process establishing connections among participants using dialogue actions so that they can represent their own positions properly.

In Figure 3-3 (c-1) and (c-2), suppose participant C has been assigned as a *side-participant* who has not engaged with other participants for a significant time. Participant C's amount of communication traffic with the other participants is significantly less than that of the others. Here, we define "*engagement density*," which represents the amount of communication traffic. As a relevant measurement of engagement density, Katzenmaier et al. produced a measure of "*utterance density*," which takes the ratio of speech to non-speech behaviour per utterance ("a speech activity per a certain unit of time by dividing each utterance duration by the sum of previous and following pause durations") (Campbell and Scherer, 2010). While the utter-

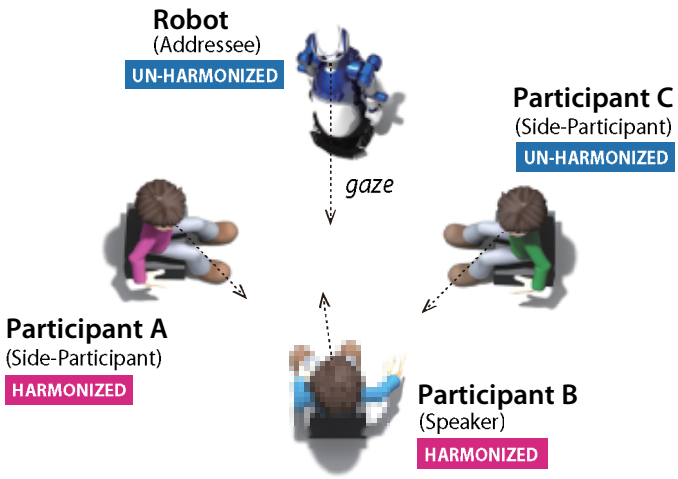


Figure 3-4: Four-participant conversational situation in our experiment. Four participants, including a robot, are talking about a certain topic. Participants A and B are leading the conversation, and mainly keep the floor. C is an *un-harmonized* participant, who does not have many chances to takemaly the floor for a while. The robot is also an un-harmonized participant at this time. The dashed arrows indicate the direction they are facing, assuming their gazes.

ance density directly dependents on speech activities, the engagement density is a measurement of amount of communication between interlocutors. Therefore, even if a participant's utterance density is high, it does not mean the engagement density is high. Jokinen et al. also mentioned that sometimes one of the participants might be less active in turn-taking (engagement) even if the speaking activity in the conversation as a whole is large (Jokinen, 2011). Three-participant conversations are likely to produce a difference of density. We define a "*harmonized*" participant as a participant with high engagement density, and an "*un-harmonized*" participant as a participant with low engagement density. Consequently, *speaker* and *addressee* are always assigned as *harmonized* participants, and *side-participants* can be divided into two types in terms of engagement density: *harmonized side-participant* and *un-harmonized side-participant*. Figure 3-3 shows the extended participation structure based on Clark's model. Although all *side-participants* are ratified, an *un-harmonized side-participant*, who is only recognized by the *speaker*, can sometimes emerge in four-participant situations.

3.2.3 Procedures Obtaining Initiatives Controlling Engagement Density

In order that a facilitator is transferred an initiative by the current speaker, the facilitator must take procedural steps. First, the facilitator must participate in the current dominant conversation the speaker is leading, try to be "harmonized" to claim an initiative, and then wait for either explicit or implicit approval from the speaker. Let us take the example shown in Figure 6-1. In the figure, participants A and B are primarily leading the current conversation. Participant C cannot get the floor to speak, and so the robot desires to give the floor to C. If the robot who is an "un-harmonized" participant speaks to C directly, without being aware of A and B, the conversation might be broken, or separated into two (A-B and C-robot), at best. In order not to break the situation, the robot should participate in the dominant conversation between A and B first, and set the stage such that the robot is approved to initiate the next situation as "harmonized" participant. According to our extended participation structure model in Figure 3-3, every person participating in

Table 3.2: Permission relationship between subject and target participants for the constraint of addressing. A "subject" means a participant who is initializing a new dialogue action to a "target" participant. "*Harmonized*" means a participant is assigned as a speaker or an addressee or a side-participant, who is harmonized with the conversational group. "*Un-Harmonized*" means a participant is assigned as an un-harmonized side-participant.

Subject \ Target	*Harmonized*	*Un-Harmonized*
Harmonized	permitted	permitted
Un-Harmonized	permitted	NOT permitted

Table 3.3: Permission relationship for permission between subject and target participants in the constraint of topic shifting.

Subject \ Target	*Harmonized*	*Un-Harmonized*
Harmonized	permitted	NOT permitted
Un-Harmonized	NOT permitted	NOT permitted

a dominant conversation is at "*harmonized*" state (participant A, B in Figure 6-1), and the other is at "*un-harmonized*" state (participant C and a robot). After participating in the dominant conversation between A and B, the robot is approved as a "harmonized participant" to initiate the conversation.

In terms of the way of controlling engagement, Whittaker et al. analyzed two-participant dialogues to investigate the mechanism how each control was signaled by speakers and how it affects discourse structure, including the lower control level, topic level and global organization level (Whittaker and Stenton, 1988). For the control level, they found that three types of utterances (prompts, repetitions and summaries) were consistently used to signal. For the topic level, they found that interruptions introduce a new topic. And the global organization is organized also by topic initiation. This study argued that not only signal utterances but also topic shifting/initialization plays an important role for engagement control. On the basis of these discussions above, we define the following constraints for both *harmonized* and *un-harmonized* participants when they address a next speaker and shift current topics:

1. **Constraint of addressing**:
   An un-harmonized participant must not address the other un-harmonized participants directly.

2. **Constraint of topic shifting**:
   A harmonized participant must not shift the current topic when he/she addresses the other un-harmonized participants.

The relationship between subject and target participants that are permitted to approach in the two constraints are shown in Tables 3.2 and 3.3. For examples, while a *harmonized* participant (speaker, addressee and harmonized side-participant) can address an both *harmonized* (addressee and harmonized side-participant) and *un-harmonized* (un-harmonized side-participant) participants, an *un-harmonized* participant can not address another *un-harmonized* participant. In the following sections, we describe a computational model that has the group maintenance functions discussed above.

Table 3.4: Samples of adjacency pairs

Adjacency Pair	Example
greeting → greeting	”Heya!” → ”Oh, hi!”
offer → acceptance/rejection	”Would you like to visit the museum with me this evening?” → ”I’d love to!”
request → acceptance/rejection	”Is it OK if I borrow this book?” → ”I’d rather you didn’t, it’s due back at the library tomorrow”
question → answer	”What does this big red button do?” to ”It causes two-thirds of the universe to implode”
complaint → excuse/remedy	”It’s awfully cold in here” → ”Oh, sorry, I’ll close the window”
degreeting → degreeting	”See you!” → ”Yeah, see you later!”
inform → acknowledge	”Your phone is over there” → ”I know”

3.2.4 Adjacency Pairs : Timing of Initializing a Procedure

In order to detect timing of initializing a procedure, a facilitator should care about a unit of consecutive sequence to avoid to break a current conversation. An adjacency pair is a minimal unit of conversational sequence organization (Schegloff and Sacks, 1973), therefore it might be reasonable to employ here. An adjacency pair is characterized by certain features (Schegloff, 2007): a) composed of two turns, b) by different speakers, c) adjacently placed, d) these two turns are relatively ordered; that is, they are differentiated into “first part parts” and “second pair parts”. First pair parts are utterance types that initiate some exchange, such as question, request, offer, invitation, announcement, etc. Second pair parts are utterance types that are responsive to the action of prior turn, such as answer, grant, reject, accept, decline, agree/disagree, acknowledgement, etc. e) pair-type related; that is, not every second pair part can properly follow any first pair part. Adjacency pairs compose pair types; types are exchanges, such as greeting-greeting, question-answer, offer-accept/decline, and the like. To compose an adjacency pair, the first and second pair parts come from the same pair type. Table 3.4 shows samples of adjacency pairs.

The basic practice or rule of operation, then by which the minimal form of the adjacency pair is produced is: 1) given the recognizable production of a first pair part, 2) on its first possible completion its speaker should stop, 3) a next speaker should start (often someone selected as next speaker by the first pair part), and 4) should produce a second pair part of the same pair type. Adjacency pair-based sequences can come to have more than two turns. Schegloff discussed expansions of adjacency pairs, including pre-expansion, insert expansion, and post-expansion. The pre-expansion comes before the first pair part. Examples of pre-expansions include pre-invitation, pre-offer, pre-announcement and other pre-telling, pre-sequence, such as summons-answer sequences that usually occurs in phone calls. The insert expansion is one that happens between a first pair part and a second pair part. Examples of the insert expansion include post-first insert expansions, pre-second insert expansions, and expansions of expansions. A minimum post-expansion happens after a second pair part as a sequence-closing third. Sequence-closing thirds takes a number of forms or combinations of them, three of the most common are “oh,” “okey,” and assessments. “Oh” registers a just-preceding utterance as an informing, as producing a change in its recipient from non-knowing to now-knowing. “Okey” (and some variants, such as “alright”) marks or claims acceptance of a second pair part and the stance that is has adopted and embodies within the sequence. An Assessment in third position articulates a stance taken up, ordinarily by the first pair part speaker, toward what the second pair part speaker has said or done in the prior turn. The product of these features of adjacency pairs may be represented schematically in a very simple transcript diagram as follows:

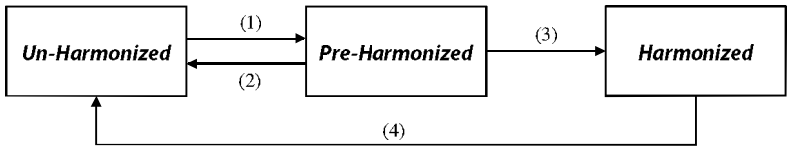


Figure 3-5: Transition of harmony states. (1) A participant claims an initiative with a first pair part, against a current speaker who is leading the current dominant conversation, waiting for either explicit or implicit approval by the speaker's second pair part. (2) A claim was declined by the speaker either explicitly or implicitly. (3) A claim was approved by the speaker's second pair part addressed to the participant who claimed an initiative. (4) An "*harmonized*" state is gradually falling down to "*un-harmonized*" while a participant is assigned as a side-participant.

←*Pre-Expansion*
A: First Pair Part
←*Insert Expansion*
B: Second Pair Part
←*Post-Expansion (sequence-closing third)*

So, which timing can be candidates for a facilitator to initiate procedures? As a facilitator might produce economically short steps of procedures to help a left behind participant, in this paper, we assume every second or third part might be the candidates to initiate. Figure 3-5 shows transition of harmony state, which describes how a facilitator makes himself/herself harmonized and takes an initiative to control a situation, by employing a concept of adjacency pairs. We assume that an *un-harmonized* participant needs to be approved by a speaker's second pair part to be harmonized. In the following sections, we will describe a computational model of the procedural process discussed above.

3.3 Engagement Density Control Procedure Optimization as POMDP

In this section, we discuss and present a computational model to enable procedures controlling engagement density. We summarized three procedural steps in Section 3.2.1: **Observation**, **Obtaining an initiative**, and **Floor and topic maintenance**. Since the model needs such a procedural decision making process, we employ Markov decision process (POMDP) (Williams and Young, 2007), which can maintain parallel state hypotheses and confidence scoring, and cope better with observation errors. In the next subsections, at first, we describe the POMDP basics, and extend it to four-participant group maintenance model.

3.3.1 Partially Observable Markov Decision Process (POMDP) Basics

In general, formulation of a POMDP can be defined as the following components: $\beta = \{S, A, T, R, O, Z, \eta, b_0\}$, where S represents a set of states of the agent's world, A represents a set of actions of the agent, T represents a transition probability $P(s'|s, a)$, R represents the instant expected reward $r(s, a)$, O represents a set of observations the agent can receive, and Z represents an observation probability, $P(o'|s', a)$, η represents a discount factor ($0 < \eta < 1$), and b_0 represents an initial belief state $b_0(s)$. In the POMDP framework, since s is a partially observable state, a distribution of states is defined as belief state b. At each time-step, the agent selects an action $a \in A$ based on b, then receives a reward $r(s, a)$. The transitions probability to a next state s' depending only on s and a. The agent receives an observation $o' \in O$ depending on s' and a. b

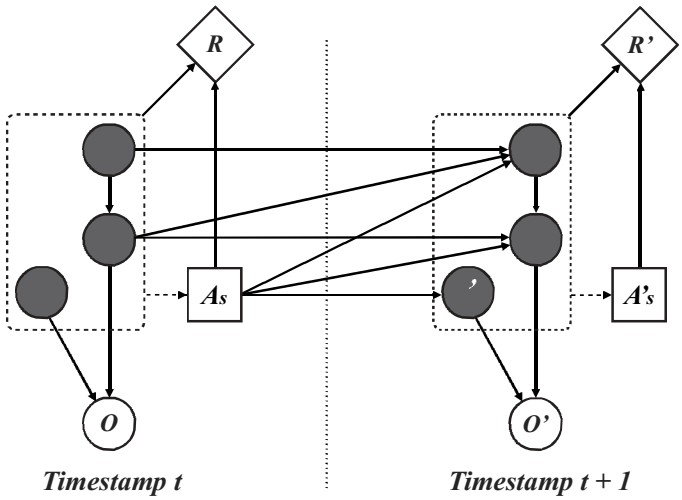


Figure 3-6: Influence diagram representing the proposed POMDP model. Circles, squares and diamonds represent random variables, decision nodes and reward nodes respectively. Shaded circles represents random variables and unshaded circles represent observed variables.

can be updated as follows:

$$\begin{aligned} b'(s') =& p(s'|o', a, b) = \frac{p(o'|s', a, b)p(s'|a, b)}{p(o'|a, b)} = \frac{p(o'|s', a)\sum_{s\in S} p(s'|a, b, s)p(s|a, b)}{p(o'|a, b)} \\ =& \frac{p(o'|s', a)\sum_{s\in S} p(s'|a, s)b(s)}{p(o'|a, b)} \end{aligned} \tag{3.1}$$

The numerator includes the observation function, transition matrix and current belief state. Since the denominator is independent of s', it can be regarded as a normalization constant η. Therefore the belief state update can be:

$$b'(s') = \eta \cdot P(o'|s', a) \sum_{s} P(s'|s, a)b(s) \tag{3.2}$$

At each state, the agent selects an action and receives reward r_t. The return, cumulative discounted reward, is given by:

$$R = \sum_{t=0}^{\infty} \lambda^t r_t \tag{3.3}$$

where λ is the discount factor $0 < \lambda < 1$. A policy π maps from belief state to action $\pi(b) \in A$, and an optimal policy $\pi^*(b) \in A$ is a policy that maximizes the expected return $E[\lambda]$.

3.3.2 Four-Participant Group Maintenance Model

In order to realize the three-step group maintenance procedure, we define the states, system actions and rewards of the extended POMDP. As the first step (observation), it is essential to know the existence of un-harmonized (left behind) participant in a current time, which we defined as an extension of the Goffman

Table 3.5: Robot's harmony states s_h

Harmony states	Meaning
Un-Harmonized	The robot is not harmonized with the current conversation.
Pre-Harmonized	The robot is waiting for approval to harmonize with the current conversation.
Harmonized	The robot is harmonizing with the current conversation.

and Clark's participation structure in Section 3.2.2. As the second step (obtaining an initiative) and third step (floor and topic maintenance), the system should obey the constraints of addressing and topic shifting we discussed in Section 3.2.3. And the procedure initiation timing can be defined by employing adjacency pairs, as we discussed in Section 3.2.4. Also, in order to manage the topic shifting, the system should know the un-harmonized participant's motivation to talk about a current topic.

Based on these considerations, we reasonably defined observations as follows: a current status of harmony (a current un-harmonized participant's ID and the robot's own harmony status), and an un-harmonized participant's motivation to speak about a current topic, and a current adjacency pair part to decide if it's allowed to initiate or continue a procedure. Such partially observable information would be given by external modules outside POMDP module (the whole architecture will be described in Section 3.4), and they could be assumed to have the Markov property. dialogue actions giving a floor to an un-harmonized participant, which would be divided into distinctive two types of actions: initiating a new topic and maintaining a current topic. And the constraints of the procedure we assumed in Section 3.2.3 (constraints of addressing and topic shifting), can be given as rewards in POMDP.

Now, we assume a set of states S can be factored into three components: the harmony states s_h, the participants' motivation states S_m, and the participants' actions A_p. Hence, the factored POMDP state S is defined as:

$$s = (s_h, s_m, a_p) \tag{3.4}$$

and the belief state b becomes as follows:

$$b = b(s_h, s_m, a_p) \tag{3.5}$$

To compute the transition function and observation function, a few intuitive assumptions are made:

$$\begin{aligned} P(s'|s,a) =& P(s'_h, s'_m, a'_p|s_h, s_m, a_p, a_s) \\ =& P(s'_h|s_h, s_m, a_p, a_s)\cdot \\ & P(s'_m|s'_h, s_h, s_m, a_p, a_s)\cdot \\ & P(a'_p|s'_m, s'_h, s_h, s_m, a_p, a_s) \end{aligned} \tag{3.6}$$

Figure 3-6 shows the influence diagram depiction of our proposed model. We assume conditional independence as follows.

3.3.3 Harmony Model

The first term in (3.6), which we call the *harmony model* T_{s_h}, indicates how participants harmonize in the current dominant conversation at each time-step. We assume that the participants' harmony state at each time-step depends only on the previous harmony state, the participants' action, and the system action. The

Table 3.6: Un-harmonized participant's motivation states S_m

Motivation states	Meaning
Motivated	The participant who is left behind has a motivation to speak on the current topic (interested in the current topic).
Not-Motivated	The participant who is left behind does not have any motivation to speak (not interested in the current topic).
none	Nobody is left behind.

transition probability can be described as follows:

$$T_{s_h} = P(s'_h|s_h, a_p, a_s) \tag{3.7}$$

Table 3.5 shows the states of harmony. In this paper, the *harmony model* only contains the robot's harmony states. In a four-participant group situation including a robot, as a speaker and an addressee are automatically assigned to be harmonized based on our definition, an un-harmonized participant exists at most only one at same time except for a robot (in Section 6-1, only participant C is an un-harmonized participant). Because the determined current participation roles (speaker/ addressee/ harmonized side-participant/ un-harmonized side-participant) are given by the role estimation module that will be described in Section 3.4.1, s_h only has to estimate the robot's harmony state in this four-participant model. The probabilities of (3.7) were handcrafted, based on the consideration in Section 3.2.3 and 3.2.4. As Figure 3-5 shows, when the harmony state is the *Un-Harmonized* state and the robot is asked by a current speaker, the state should be changed to the *Pre-Harmonized* state, where the robot is awaiting the speaker's approval for the *Harmonized* state. We assume that any dialogue acts from the speaker addressing the robot in the *Pre-Harmonized* are approvals. Otherwise, the state will be back to the *Un-Harmonized*. The *Harmonized* state gradually goes down to the *Un-Harmonized* state in time-steps unless the robot selects any dialogue acts.

3.3.4 Motivation Model

We call the second term the *participants' motivation model* T_{S_m}, which indicates how an un-harmonized participant has the motivation to take the floor at each time-step. This state implies that the participant who is left behind (target person) has a motivation to speak on the current topic. Thus, this state affects decision-making about topic maintenance. Estimated un-harmonized participants' motivation at each time-step is given by the motivation estimation module that will be described in Section 3.4.2. And we assume that a participant's motivation also depends on the previous system action. The transition probability can be described as follows:

$$T_{S_m} = P(s'_m|a_s) \tag{3.8}$$

Table 3.6 shows the left behind participant's motivation states.

3.3.5 Participants' Action Model

We call the third term the *participants' action model* T_{A_p}, which indicates what actions the participants are likely to take. We assume the participants' action at each time-step depends on the previous participant's action, the previous system action, and the current robot's harmony state. The transition probability can be

Table 3.7: Participants' actions A_p

Participants' actions	Meaning
first-part	A participant made a first pair part
second-part	A participant made a second pair part
third-part	A participant made a sequence closing third
first-part-toRobot	A participant made a first pair part to a robot
second-part-toRobot	A participant made a second pair part to a robot
third-part-toRobot	A participant made a sequence closing third to a robot

Table 3.8: System actions A_s

System actions	Meaning
answer	Answering a current speaker's question
question-new-topic	Asking someone a question related to a new topic
question-current-topic	Asking someone a question related to a current topic
opinion	Giving own opinion or a trivia
simple-reaction	Reacting to a current speaker's call of robot's name.
nod	Nodding to a current speaker
none	Doing nothing but giving a gaze to a current speaker

described as follows:

$$T_{A_p} = P(a'_p|s'_h, a_p, a_s) \tag{3.9}$$

Participants' actions are defined as adjacency pairs as shown in Table 3.7. As we discussed in Section 3.2.4, understanding adjacency pairs, minimal units of conversational sequences, is essential to detecting a timing of initializing a procedure. We assume recognizing three parts (first/second/third) is sufficient to detect the timing. Pair types (e.g. greeting-greeting, question-answer, offer-accept/decline) are not distinguished in this case. The transition probabilities of adjacency pair types are based on a corpus we collected. We recorded two four-participant conversational groups (all participants were human subjects), who were given the task of discussing movies. The total duration was around 60 minutes. Each utterance is segmented automatically by our speech recognition. After the recording, adjacency pair types were manually annotated for all speech segments.

3.3.6 System Actions

Table 3.8 shows the system actions. The system has six actions available. *Answer* action is answering a current speaker's question, triggering the Answer Generation module though the Content Planner module. The question action is divided into two types: *question-new-topic* and *question-current-topic*. *Question-new-topic* is a question action with initiating a new topic. A robot can use this action according to the constraint of topic shifting as we discussed in Section 3.2.3 (Table 3.3). *Question-current-topic* is a question action along a current topic, without topic shifting. *Simple-reaction* is a simple reacting action to a current speaker's call of robot's name (e.g."SCHEMA!"). *Nod* generates a nod to a current speaker in other to indicate that the robot is listening to a current speaker's utterance. When the robot has not an initiative controlling a situation, it is most likely to select this action to avoid breaking a current conversational sequence. *None* does nothing but giving a gaze to a current speaker. Both *nod* and *none* are also likely to be used when a belief is not high enough to select a dialogue action.

Table 3.9: Examples of rewards r associated with a timing of initializing a procedure. A left-behind participant has already detected. "*" represents any states.

Harmony state s_h	Participants' actions a_p	Motivation state s_m	System actions a_s	Rewards r
*	*1st* to robot	*	*answer*	+5
*	*2nd* or *3rd* to robot	*	*nod*	+5
Un-Harmonized	*1st* to other	*	*opinion*	+4
Un-Harmonized	*1st* to other	*	*nod*	+3
Un-Harmonized	*2nd* to other	*	*opinion*	+3
Un-Harmonized	*2nd* to other	*	*nod*	+3
Pre-Harmonized	any utterance to robot	*Motivated*	*question-current-topic*	+5
Pre-Harmonized	any utterance to robot	*Not-Motivated*	*question-new-topic*	+5
Pre-Harmonized	any utterance to robot	*	*nod*	+5
Harmonized	*2nd* or *3rd* to robot	*	*question-current-topic*	+5
Harmonized	*2nd* or *3rd* to robot	*	*question-new-topic*	-5

3.3.7 Belief State Update

We define the observation probability Z as follows:

$$Z = P(o'|s', a) = P(o'|s'_m, a'_p, a_s) \tag{3.10}$$

Given the definitions above, the belief state can be updated at each time-step by substituting (3.7), (3.8), and (3.9) into (3.2):

$$\begin{aligned} b'(s'_m, a'_p) = \eta \cdot & \underbrace{P(o'|s'_m, a'_p, a_s)}_{\substack{\text{observation} \\ \text{model}}} \cdot \sum_{s_m} \underbrace{P(s'_m|a_s)}_{\substack{\text{motivation} \\ \text{model}}} \cdot \sum_{a_p} \underbrace{P(a'_p|s'_h, a_p, a_s)}_{\substack{\text{participants'} \\ \text{action model}}} \cdot \\ & \sum_{s_h} \underbrace{P(s'_h|s_h, a_p, a_s)}_{\substack{\text{harmony} \\ \text{model}}} \cdot b(s_m, a_p) \end{aligned} \tag{3.11}$$

On the basis of the consideration of the constraints in Section 3.2.3, the reward measure includes components for both the appropriateness and inappropriateness of the robot's behaviors. Table 3.9 shows examples of rewards we used in our experiments.

As an optimization algorithm, we employed approximate value iteration methods with point-based updates. These algorithms have proven to scale very effectively, relying on the fact that performing many fast approximate updates often results in a more useful value function than performing a few exact updates. In this paper, we employed the heuristic search value iteration (HSVI) algorithm proposed by Smith et al., which is one of point-based algorithms (Smith and Simmons, 2012). We used ZMDP[1] as a policy optimization tool.

[1] http://www.cs.cmu.edu/ trey/zmdp/

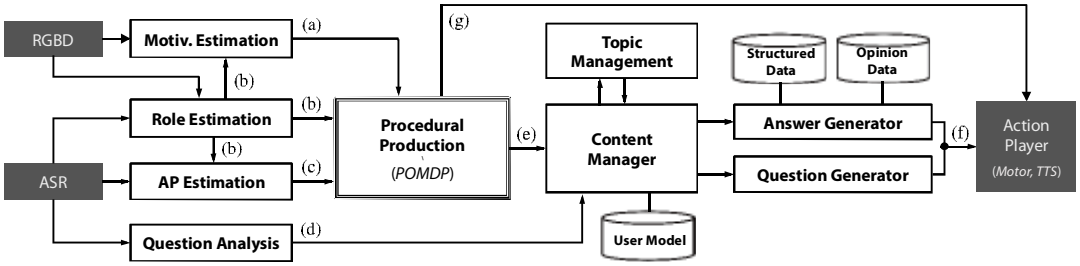


Figure 3-7: The architecture of the system primarily comprises the situation understanding process (Participation Role Recognition, Adjacency Pair Recognition and Motivation Estimation), the POMDP based procedural production process described in Section 3.3, and the language generation process (Question Analysis, Content Planning, Topic Management, Answer Generation and Question Generation). The situation understanding process receives sensory information from RGBD cameras (Microsoft Kinect) and automatic speech recognizers (ASR) for each participant. Action Player consists of Motor Control and Text to Speech modules. (a)-(g) represent each output from each module: (a) a left behind participant's motivation, (b) estimated roles including harmonized/un-harmonized side-participant, (c) estimated an adjacency pair part, (d) interpreted question types, (e) determined a system action, (f) a generated sentence and its target person ID to be addressed, and (g) gaze control information (target person ID) transmitting to Action Player interpreting as a concrete position.

3.4 System Architecture

Based on the studies on small group maintenance, we propose an architecture for conversational robots that has the capability to facilitate small groups, as shown in Figure 4-4. The framework primarily comprises three processes: situation understanding, procedural production and language generation. The situation understanding process consists of Participation Role Recognition, Adjacency Pair Recognition, and Motivation Estimation. The procedural production process produces procedural actions maintaining a small group, based on the POMDP model we described in Section 3.3. The language generation process consists of Question Analysis, Content Planning, Topic Management, Answer Generation and Question Generation. Each participant has a wireless microphone on its chest, connected to each Automatic Speech Recognizer (ASR). RGBD cameras are also set in front of each participant. Each time the system detects a voice activity detection (VAD) by each participant's ASR module, the procedural production is triggered to process information interpreted by the situation understanding process. Content Planner generates a concrete sentence, as referring a current topic and user models. It calls either Answer Generation or Question Generation according to a determined dialogue action output from Procedural Production. In the following subsections, we describe each module of the situation understanding and the language generation processes.

3.4.1 Participation Role Recognition

The role estimation module manages participation roles presented in Figure 3-3. In this paper, we employ the following assumptions for role classification in a four-participant situation.

1. One *speaker* always exists in one group at each time-step.
2. One *addressee* who is addressed by the *speaker* always exists at each time-step.
3. A *side-participant* is a participant who is not assigned neither *speaker* nor *addressee*.

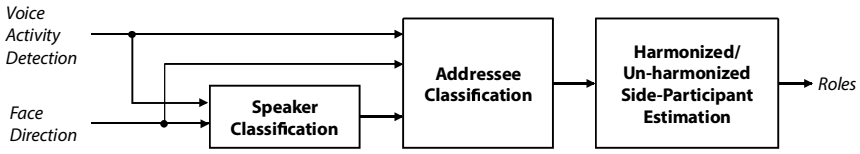


Figure 3-8: Participation role recognition process. Participation roles including a speaker, an addressee, and side-participants are recognized by the results of voice activity detection (VAD) and face directions recognition. Speaker classification is based on results of face direction classification and VAD. Addressee classification is based on a result of speaker classification, as well as face direction and VAD. As the final process, a side-participant is classified either "*Harmonized*" or "*Un-Harmonized*" The face directions are captured by depth-RGB cameras (Microsoft Kinect).

As we defined in Section 3.2.2, *side-participants* can be divided into two types: *harmonized side-participant* and *un-harmonized side-participant*. Figure 3-8 shows the role-estimation process consisting of distinctive three sub-processes: speaker classification, addressee classification, and harmonized/un-harmonized side-participant estimation. The speaker classification is based on the results of face direction classification and VAD. The addressee classification is based on the result of speaker classification, as well as each participant's face direction and VAD. The face directions are captured by depth-RGB cameras (Microsoft Kinect). The best results of classification using Naive Bayes for speaker and addressee classification were 79.4% and 70.9%, respectively.

In the final process, another participant, who should be assigned to a side-participant according to our definition above, is estimated whether he/she is harmonized or un-harmonized. In the scenario shown in Figure 6-1, participant C may not be able to take the floor for a while. We assume the situation probably resolves itself when the current topic is shifted. Hence, we define the depth of side-participant $Depth_{SPT}$ as the duration that a participant is assigned while the same topic continues, which represents the level of harmony.

$$Depth_{SPT_i} = Duration_{SPT_i} / Duration_{topic_j} \tag{3.12}$$

$$HarmonizedSPT = \begin{cases} SPT_i & \text{if } Depth_{SPT_i} > Threshold \\ none & \text{otherwise} \end{cases} \tag{3.13}$$

where the suffix i represents a participant's ID.

3.4.2 Motivation Estimation

As we discussed in Section 3.3.4, the motivation estimation manages only an un-harmonized participant's motivation to take a floor on the current topic. Thus, this state affects decision making about topic maintenance. We define motivation as an un-harmonized participant's ID and a binary (true/false) variable, which is heuristically calculated as follows:

$$Motivation_i = \begin{cases} 1 & \text{if } MotivationAmount_i > Threshold \\ 0 & \text{otherwise} \end{cases} \tag{3.14}$$

In our previous experiment, we analyzed how a conversational robot's existence and its actions can affect users' impressions in group game situations, using video analysis, SD (Semantic Differential) method and free-form questionnaires. The result of SD method indicates that subjects feel more pleased, and the re-

Table 3.10: Speaker estimation accuracy using Naive Bayes [%]

Feature	estimation accuracy
VAD only	75.1
Gaze pattern only	56.6
VAD + Gaze pattern	**79.4**

Table 3.11: Addressee estimation accuracy using Naive Bayes [%]

Feature	estimation accuracy
Result of speaker classification + VAD	36.0
Result of speaker classification + Gaze pattern	**70.9**
Result of speaker classification + VAD + Gaze pattern	68.3
Result of speaker classification + Speaker's gaze direction	66.1
Result of speaker classification + VAD + Speaker's gaze direction	67.2

sults of free-form questionnaires showed many participants were motivated to participate in the game, with participation and active actions of a robot. These psychological results correlate with utterance frequency and smiling duration ratio, calculated by annotated data (Matsuyama et al., 2010). Also, according to our observation and discussions of the experiments, even if participant's utterances are not observed frequently, participants motivated to participate are likely to nod frequently, as reacted to a speaker's utterances. Therefore, we assume the amount of motivation of a participant can be calculated by a heuristic linear function of speech, smiling and nodding activities during duration of a certain topic, as follows:

$$MotivationAmount_i = \int_{t_{start}}^{t_{end}} (\alpha f_{speech_i}(t) + \beta f_{smile_i}(t) + \gamma f_{nod_i}(t))dt \tag{3.15}$$

where t represents a current time. t_{start} and t_{start} represent start and end times of a continuum topic, respectively. α, β and γ are arbitrary coefficients. The speech activities are calculated using results of VAD. The smiling and nodding activities are calculated by smiling detection and nodding detection modules, using Microsoft Kinect's Face Tracking SDK[2].

3.4.3 Adjacency Pairs Estimation

In this paper, adjacency pairs are recognized by the results of participation role recognition and speech recognition. Each time the system detects an endpoint of speech from the automatic speech recognition module, it classifies each utterance into one of the six categories shown in Table 3.7 ($\{$1st, 2nd, 3rd$\}$ $\times$ $\{$toRobot, notToRobot$\}$). In this paper, adjacency pairs are recognized by the linear-chain conditional random fields (CRF), using results of speech recognition. The following features are used in the prediction process:

$$word_{t-2}, word_{t-1}, word_t, word_{t+1}, word_{t+2}, word_{t-1}\&word_t, word_t\&word_{t+1}$$
$$pos_{t-2}, pos_{t-1}, pos_t, pos_{t+1}, pos_{t+2}, pos_{t-1}\&pos_t, pos_t\&pos_{t+1}$$
$$spos_{t-2}, spos_{t-1}, spos_t, spos_{t+1}, spos_{t+2}, spos_{t-1}\&spos_t, spos_t\&spos_{t+1}$$
$$spk_{t-2}, spk_{t-1}, spk_t, spk_{t+1}, spk_{t+2}, spk_{t-1}\&spk_t, spk_t\&spk_{t+1}$$

where $word_t$, pos_t, $spos_t$ and spk_t denotes word, part of speech, subparts of speech, speaker id at time t,

[2]http://www.microsoft.com/en-us/kinectforwindows/

Table 3.12: Example of features of adjacency pair. In this example, person A initiates a first pair part ("Do you know the story of the movie?"), and person B replies to it by a second pair part ("I do not know much about it."). The BIO column represents classified BIO encoding. "B-1" and "B-2" represent beginning of a first and a second pair parts, and "I-1" 'and I-2" represent they are in a first and a second pair parts.

word (pronounciation)	part of speech	subparts of speech	speaker id	BIO
映画 (eiga)	noun	general noun	A	B-1
の (no)	particle	adverbial particles	A	I-1
内容 (naiyou)	noun	general noun	A	I-1
は (wa)	particle	adverbial particles	A	I-1
知って (shitte)	verb	verb	A	I-1
ます (masu)	postfix	verbal suffix	A	I-1
か (ka)	particle	conjunctive particles	A	I-1
あんまり (anmari)	particle	particle	B	B-2
知ら (shira)	verb	verb	B	I-2
ない (nai)	postfix	adjective suffix	B	I-2
です (desu)	verbal auxiliary	verbal auxiliary	B	I-2

respectively. Table 3.12 shows an example of features of adjacency pair we used. We use CRF++ toolkit[3] in our experiments.

For learning and evaluation, we recorded conversational data where 3 participants are assigned to each group and talked for 10 minutes. We had totally 7 groups (70 minutes with 21 participant). They were instructed that they would talk about movies within movie-related 100 topics we defined beforehand. We used 6 groups for learning, 1 group for evaluation. After we transcribed the recorded conversations, each utterance separated manually by an experimenter. Then each of them is analyzed by a Japanese language morphological analyzer[4]. The analyzer allows the part of speech to be further sub-classified, namely the subparts of speech. Based on the analyzed results, we coded each morpheme with an extended BIO encoding scheme. Using the BIO, each word is tagged as either (B)eginning an entity, being (I)n an entity, or being (O)utside of an entity. In this case, we extended it with adjacency pairs: a beginning of a first pair part is coded as "B-1", and subsequent words are coded as "I-1." The same rule is applied for both second and third parts ("B-2" or "I-2" for second parts, "B-3" or "I-3" for third parts). As for the successfulness of the coding, the inter-rater agreement using Cohen's kappa (Fleiss et al., 2013) indicated a substantial result between the two raters ($\kappa = 0.75$). The classification accuracy for each word was 73.5%. And a result of the last word will be the final result of the adjacency pair.

3.4.4 Topic Management

In this paper, we define a sequence of topic words as a conversational context. Each system utterance is hooked to one of the topics. For example, the sentence "*Audrey is beautiful, isn't she?*" is assumed to belong to the topic "Audrey Hepburn." In our experimental system, we prepared 100 topic words for each domain. The topics in the *movies* domain include genres, titles, directors, and actors.

The topic estimation procedure uses the following three processes: Japanese language morphological analysis, important words filtering, and classification. After an ASR or text input is processed by Japanese language morphological analysis, only nouns are extracted. Then, the important nouns in each topic are

[3]https://code.google.com/p/crfpp/
[4]http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN

extracted. In terms of degrees of importance, we use the term frequency-inverse document frequency (TF-IDF) score, which is often used as a weighting factor in information retrieval and text mining. We collected the top 64 web sites as 64 separate documents for each topic word using Google web search. In the classification process, we used the linear-chain conditional random fields (CRF) technique. We use CRF++ toolkit[5] in our experiments. The following features are used in the prediction process:

$$topic_{t-2}, topic_{t-1}, topic_t, topic_{t+1}, topic_{t+2},$$
$$topic_{t-2}\&topic_{t-1}\&topic_t, topic_{t-1}\&topic_t\&topic_{t+1}, topic_t\&topic_{t+1}\&topic_{t+2}$$

where $topic_t$ denotes the topic word at time t. As an evaluation experiment, we evaluated the accuracy of 10-topic classification. We recorded three-minute conversations with two participants in which they were instructed to talk within 10 topics in the animation film domain. We conducted a total of 25 sessions. 20 of which were used for learning data, and five used for test data. One experimenter annotated each word as correct answers. The result for the accuracy rate (number of correct answers / total number estimated) was 88.2% under a word error rate for ASR of 0%, and 64.7% under a word error rate for ASR of 20%.

3.4.5 Question Generation

The Question Generation Module has two main functions: giving someone the floor and collecting the user model. The user model is preferred for topic maintenance. We define that a user's interests in a certain topic are organized by experiences and preferences. The system extracts this information in the following ways.

1. User's answer to the system's question.
   The system directly asks a user his/her experiences and preferences about a certain topic.

2. User's motivation (interests) for each topic.
   When a topic transition occurs, the system obtains each user's preference, which is calculated as the sum of their motivation during the topic.

A preferred new topic is determined using cosine similarity of TF-IDF scores. The topic scores ($TopicScore$) of all topics are calculated on the basis of the cosine similarities of the current topic ($CurrentTopic$), a user's topic preferences of all topics ($PreferenceTopic$), and experiences ($ExperienceTopic$) between the $CurrentTopic$ and each $Topic$.

$$\begin{aligned} TopicScore_i = {} & \alpha \cos(Topic_i \cdot CurrentTopic) \\ & + \beta \left(\sum_m \cos(Topic_i \cdot PreferenceTopic_m) \right) \\ & + \gamma \left(\sum_n \cos(Topic_i \cdot ExperienceTopic_m) \right) \end{aligned} \tag{3.16}$$

where $\alpha > \beta > \gamma$. According to the $TopicScore$, the system can shift a topic to another that is close to a left-behind participant's interest.

3.4.6 Answer Generation

Based on the results of the Question Analysis process, answers are classified into two types: Factoid type answers and Non-factoid type answers (opinions). Factoid answers are generated from a structured database.

[5] https://code.google.com/p/crfpp/

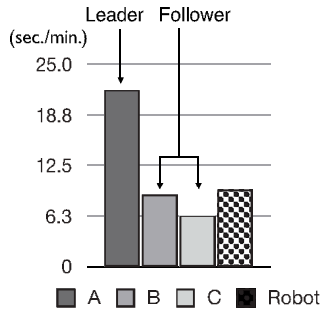


Figure 3-9: Leader and follower

In this research, we use Semantic Web technologies. After analyzing a question, it is interpreted as a SPARQL query, a resource description framework (RDF) format query language to search RDF databases. We use DBpedia as an RDF database[6]. The opinion (non-factoid type answers) generation process refers opinion data automatically collected from a large amount of reviews in the Web. The opinion generation consists of four process: document collection, opinion extraction, sentence style conversion, and sentence ranking. As an example task, we collected review documents from the Yahoo! Japan Movie site[7]. For further explanations of the mechanisms of the Answer Generator, see (Matsuyama et al., 2014).

3.4.7 Experimental Platform

For our experimental platform, we used the multimodal conversation robot "SCHEMA([∫e:ma])," (Matsuyama et al., 2009) shown in Figure 6-1. SCHEMA is approximately 1.2[m] in height, which is the same as the level of the eyes of an adult male sitting down in a chair. It has 10 degrees of freedom for right-left eyebrows, eyelids, right-left eyes (roll and pitch) and neck (pitch and yaw). It can express anxiousness and surprise using its eyelids and control its gaze using eyes, neck, and autonomous turret. In addition, it has six degrees of freedom for each arm, which can express gestures. One degree of freedom is assigned to the mouth to indicate explicitly whether the robot is speaking or not. A computer is inside the belly to control the robot's actions, and an external computer sends commands to execute various behaviors though a WiFi network. All modules, including the ASRs and a speech synthesizer are connected to each other though a middleware called the Message-Oriented NEtworked-robot Architecture (MONEA), which we earlier produced (Nakano et al., 2006). Figure 5-3 shows an example sequence of the proposed system.

3.5 Experiments

3.5.1 Preliminary Experiment

Before the experiments evaluating the efficiency of our proposal procedures, we implemented a simple system with "naive" strategies (with out the proposal procedures) and conducted a preliminary experiment with the following two conditions to discuss the effects.

[6]http://ja.dbpedia.org/
[7]http://movies.yahoo.co.jp

- **Condition 1 (passive robot):** The baseline robot system acts passively without any proactive actions. This system has the following basic skills.
  - Topic management: In order to follow the current context, the system has topic management skills described in Section 3.4.4.
  - Participation role recognition (with utterance density measurement): While the system has a participation role recognizer, it measures utterance density, instead of engagement density described in Section 3.4.1, in order to detect a left behind participant. Here, we define "*participation barrier*" as a difference between own utterance density (e.g. person A) and a sum of the other two participants' utterance density (e.g. person B and C). If a sum of the other participants' utterance density is high and own one is low, own participation barrier tends to be higher. In the final process of the recognizer of this condition, a participant who has the higher participation barrier than a certain threshold is the next target participant.
  - Motivation estimation: The same module described in Section 3.4.2
  - Answer/Question Generator: A similar QA module described in Section 3.4.5 and 3.4.6
- **Condition 2 (proposal):** The proposed robot's ability consists of seven different skills including the three skills above. Detecting the person who had not talk for a while are additional skills for proposal robot. Detecting the motivation of the those person are proposal skill as an observation roles. Asking the question to that person are proposal skill as a floor maintenance roles. Changing the topic are also proposal skill as topic maintenance roles.

Thirty Japanese students at Waseda university were hired for this experiment (16 men and 14 women). The majors of the all students are spread among different fields including the economics, social science, education, literature, and computer science. Age ranges were between 19 and 39 and the average of the age was 22.8. The groups were divided in two groups based on the tendency of interpersonal skills. All subjects were requested to answer the questionnaire along the Kikuchi's KiSS-18 (Kikuchi, 2004). This questionnaire measures the degree of the social skill defined as "skills for facilitating the interpersonal relationship with no or less difficulties" (Yoshida and Hori, 2001). The questions were arranged for six different social skills such as fundamental skill, applied social skill, skills for dealing the emotions, skills for alternatives for attacking, skills for dealing the stress, and skills for management. All subjects were requested to choose each answer from five levels of options. Because the average of the result was 3.27, the subjects were divided among above and below the score of 3. Each group consisted of three members, and seven groups with higher score and three groups with lower score were made. Each group was asked to have a free conversations within 50 topics related with movies in each trial, and the topics were written on the paper. The topic were changed for each trial and the subjects can check the paper during the experiment. The subjects were told that the robot would participate in the conversation. Every group had two different conditions. Each condition continued about ten minutes. The order of the conditions were changed for each group. As the experimental platform, we used "SCHEMA([$\int$e:ma])."

We compared amounts of utterance and silent in two conditions, The result showed the evidence that a amount of utterance of each subject was increased with the robot's proactive behavior giving chances of taking floors.

We found that at least two types of participants spontaneously arise in any types of groups in terms of utterance density as is shown in Figure 3-9. We call a "leader," a persons who lead the conversation (we assume there is only one leader in one group), and "followers," persons who follow the leader. Average

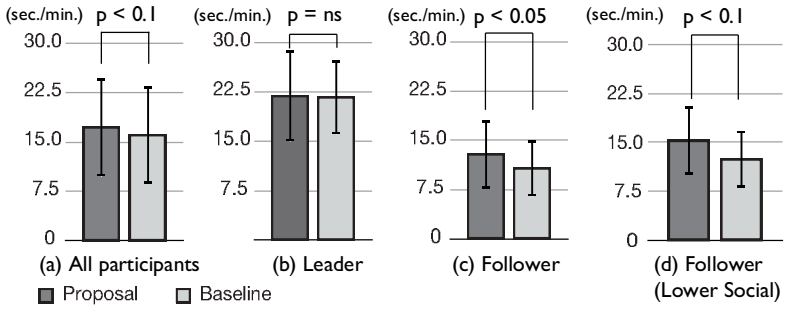


Figure 3-10: Means of duration of utterances (sec./min.)

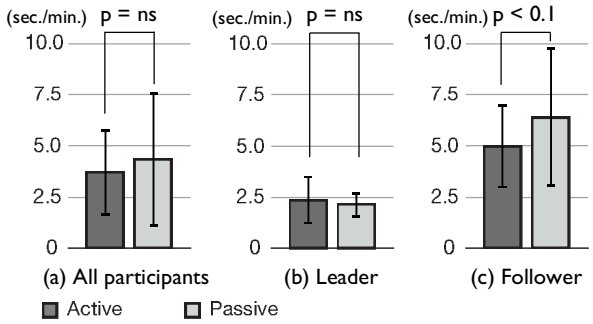


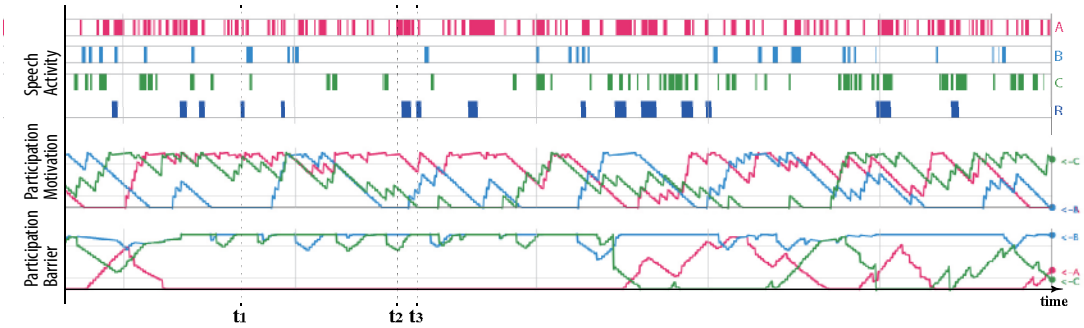
Figure 3-11: Means of duration of silences (sec./min.)

of the amount of follower's utterance in two different conditions has a significant difference ($t(13) = 2.28, p < 0.05$), as is shown in Figure 3-10 (c). Also average of the amount of follower's silent in two different conditions has a significant difference ($t(13) = -1.97, p < 0.1$), as is shown in Figure 3-11 (c).

However, while the result shows the effect of the robot's proactive behaviors, there are still two critical problems. First, the timing of initializing a robot's utterance frequently interrupted a conversation where current speaker and addressee were continuing. We had some answers of the questionnaire, such as, "There were some difficult time to talk because of the interrupt of the robot during the conversation." It might suggest a facilitator should care about timings of initializing a new utterance. Second, a topic introduced in a conversation from the robot made participants hard to follow the conversation even the topic is related the previous one. In this experiment, the robot introduced a new topic was introduced when the one participant have not talked during a short time period and he/she was not much motivated. In that moment, a participant in the conversation felt uncomfortable because the robot changed the topic too aggressively. We had some answers of the questionnaire, such as, "Too many topic change were there." The factor of the sense of incongruity might be caused by lucks of the procedures of maintaining the groups. Therefore, in the following experiments, we will discuss about procedures of maintaining the groups, including its timings.

3.5.2 Experimental Design

In order to evaluate the efficiency of our proposal procedure, especially step 2 (obtaining the initiative to control the situation and wait for approval from the others) discussed in Section 3.2.3, we designed the



t1: The participant B's participation barrier is high and motivation is low (the participant B seems not interested in the latest topic).
The robot actively asks the participant B with a general topic("animation"), "Do you like the animation film?"
Then the participant B imediately replies to the robots's question, "Yes, I like animation film."

t2: The participant A asks the robot, "Have you ever watched the SAW, SCHEMA?" (The topic is estimated "SAW")
Then the robot imediately replies to the participant A with the same topic("SAW"),
"I am really lucky to belong this community. I can't live near the murder like him. "

t3: The participant B's participation barrier is high and motivation is gradually increased (the participant B seems interested in "SAW").
So the robot asks the participant B with the same topic("SAW"), "Have you ever watched the SAW? "

Figure 3-12: Excerpt of the preliminary experiment.

following three experiments. **Experiment 1** evaluates the appropriateness and feeling of groupness as results of our proposed procedures. **Experiment 2** evaluates the appropriateness of timing of initiating procedures. **Experiment 3** compares performances of POMDP and MDP models via user simulations.

While one ideal way of evaluating a facilitation robot's procedures would be to conduct in real conversational situations (truly naive three participants participate in an experiment), it is extremely difficult to maintain quality of interactions (e.g. avoiding speech recognition errors) in all conditions to focus on evaluations of the effectiveness of the use of the group maintaining procedures. Therefore, we prepared videos of four-participant conversational situations (Human person A, B, C, and a robot), where a facilitation robot initiates procedures, or naively approaches the left behind participant C without procedural steps. The spatial arrangement was the same as that shown in Figure 5-3. Each subject was requested to watch videos from a third party. Since the experiment 1 and 2 were aimed to evaluate how the existence of our proposal procedure is effective in a group, the rules of procedures in all conditions in the experiments were hand-crafted, not POMDP model in the videos. The effectiveness of POMDP itself was evaluated in the experiment 3. All modules described in Section 3.4, including ASR and RGB-D camera sensors, were used for this experiment in order that the system could run in realtime. This way allows us to maintain process time of all modules in all conditions. The person A and B in the videos acted that they had a friendly relationship with each other, and person C acted to be coming in for the first time and be left behind in the conversation. A robot system actually reacted to actors' actions. All subjects were native Japanese speakers recruited from Waseda University campus. They were first given a brief description of the purpose and the procedure of the experiments. They were instructed that, in the videos, A and B have a friendly relationship with each other, C is left behind in the conversation, and a robot is trying to maintain the harmony of this situation. We also explained the definition of "a harmonized situation": "a situation in which all participant are given their opportunities to speak something fairly, and to share their common topics among them."

In the experiment 3, the user simulation experiment would be useful enough to prove the advantages of use of POMDP for modeling the procedural decision making, with compared with MDP.

3.5.3 Experiment 1: Appropriateness and Groupness by Usage of Procedures

The purpose of the experiment 1 was to evaluate appropriateness of the proposal procedure for group maintenance, and feeling of groupness as the result of the use of the procedure. A total of 35 subjects (23 males and 12 females) participated in this experiment. The ages of the subjects ranged between 20 and 25 years, with an average age of 20.5 years. We prepare four types of videos along the following conditions. Each video was edited to be approximately 30 seconds long. All videos started from a same topic ("*Princess Mononoke*").

- **Condition 1**: Without procedures (without topic shifting). A robot directly asks an un-harmonized participant without procedures to claim an initiative. As shown in Figure 3-13, after a sequence of interactions between A and B, which is segmented by a third adjacency pair part, a robot directly addresses C. The topic is maintained ("*Princess Mononoke*").
- **Condition 2**: With procedures (without topic shifting). A robot addresses an un-harmonized participant with procedural steps (claiming an initiative, and waiting for an approval). As is shown in Figure 3-14, after a sequence of interactions between A and B, a robot addresses A with the first pair part and waits for A's response (the second part). Then, it finishes the interaction with A, and yields the floor to C. In this case, the topic is maintained ("*Princess Mononoke*").
- **Condition 3**: Without procedures, with topic shifting. As is shown in Figure 3-15, In question #6 of Condition 1, a robot initiates a new topic ("*From Up On Poppy Hill*").
- **Condition 4**: With procedures, with topic shifting. As is shown in Figure 3-16, In question #7 of Condition 2, a robot initiates a new topic ("*From Up On Poppy Hill*").

After watching each video, the participants were asked to answer 7-scale Likert questionnaires about the (a) **appropriateness of procedures** (7 is "very appropriate," 6 is "appropriate," 5 is "rather appropriate," 4 is "not sure," 3 is "rather inappropriate," 2 is "inappropriate," and 1 is "very inappropriate"), and (b) **feeling of groupness** (7 is "very harmonized," 4' is "not sure," 3 is "rather un-harmonized," 2 is "un-harmonized," and 1 is "very un-harmonized"). In order to cancel order effects, we changed the order of the four videos for each participant. In addition, the participants were also asked to complete free-form questionnaires after watching each video.

3.5.4 Experiment 2: Appropriateness of Timing of Initiating Procedures

The purpose of experiment 2 was to evaluate appropriateness of timing of initiating procedures. A total of 32 subjects (21 males and 11 females) participated in this experiment. The ages of the subjects ranged between 20 and 25 years, with an average age of 20.5 years. After they watched the videos, they were asked to complete the questionnaires about the timing of initiating procedures (e.g., "Which video did you feel was the most appropriate?").

The following three conditions were videotaped, and the video was edited to be approximately 30 seconds long. All videos contained the same topic ("*Princess Mononoke*"). The spatial arrangement was the same as that shown in Figure 6-1. We created the following conditions:

- **Condition 1 (first part)**: Initiating a procedure just after the first adjacent pair part.
- **Condition 2 (second part)**: Initiating a procedure just after the second adjacent pair part.
- **Condition 3 (No AP)**: No consideration of adjacency pairs.

#	SPK → ADD	AP	Sentences
1	**A→B**	First	Have you ever watched "Princess Mononoke"?
2	**B→A**	Second	Yes, I have
3	**A→B**	First	Oh, you have?
4	**B→A**	Second	Yeah.
5	**A→B**	Third	I see
6	**R→C**	First	**Have you ever watched "Princess Mononoke"?** (Initializing a procedure & Floor control)
7	**C→R**	Second	Yes, I have

Figure 3-13: Transcript of condition 1 (experiment 1): Without procedures (without topic shifting).

#	SPK → ADD	AP	Sentences
1	**A→B**	First	Have you ever watched "Princess Mononoke"?
2	**B→A**	Second	Yes, I have
3	**A→B**	Third	I see.
4	**R→A**	First	**It is one of my favorite movies among Ghibri's** (Initializing a procedure)
5	**A→B**	Second	Really?
6	**B→A**	Third	Yes.
7	**R→C**	First	**Have you ever watched "Princess Mononoke"?** (Floor control)
8	**C→R**	Second	Yes, I have

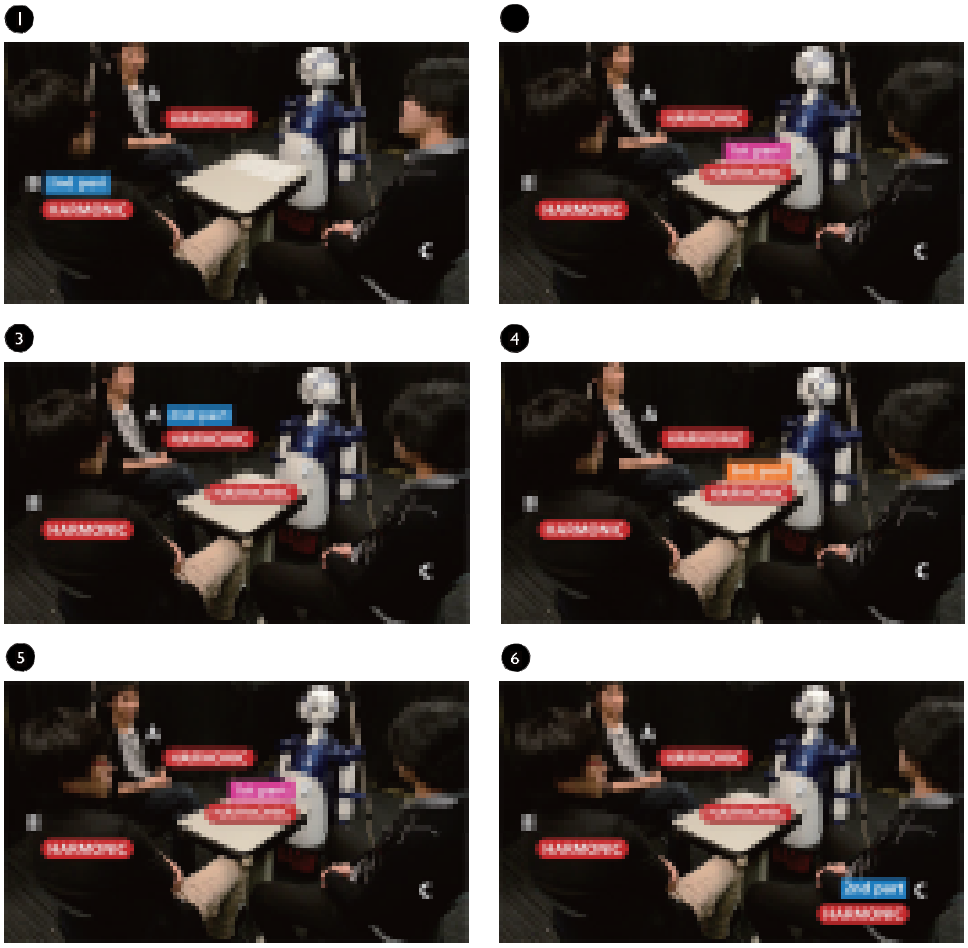
Figure 3-14: Transcript of condition 2 (experiment 1) : With procedures (without topic shifting)

#	SPK → ADD	AP	Sentences
1	**A→B**	First	Have you ever watched "Princess Mononoke"?
2	**B→A**	Second	Yes, I have
3	**A→B**	First	Oh, you have?
4	**B→A**	Second	Yeah.
5	**A→B**	Third	I see
6	**R→C**	First	**Have you ever watched "From Up On Poppy Hill"?** (Initializing a procedure & Topic shift)
7	**C→R**	Second	Yes, I have

Figure 3-15: Transcript of condition 3 (experiment 1) : Without procedures, with topic shifting

#	SPK → ADD	AP	Sentences
1	**A→B**	First	Have you ever watched "Princess Mononoke"?
2	**B→A**	Second	Yes, I have
3	**A→B**	Third	I see.
4	**R→A**	First	**It is one of my favorite movies among Ghibri's** (Initializing a procedure)
5	**A→B**	Second	Really?
6	**B→A**	Third	Yes.
7	**R→C**	First	**Have you ever watched "From Up On Poppy Hill"?** (Topic shift)
8	**C→R**	Second	Yes, I have

Figure 3-16: Transcript of condition 4 (experiment 1) : With procedures, with topic shifting



#	SPK→ADD	AP	S_h	Sentences
				(Topic: "007 Skyfall")
1	**A→B**	**1st**	*Un*	Let's talk about the "Skyfall."
2	**A→B**	**1st**	*Un*	Have you ever seen the latest one?
3	**B→A**	**2nd**	*Un*	Well, I've not seen that. ····· ❶
4	**A→B**	**3rd**	*Un*	Oh, really.
5	**R→A**	**1st**	*Pre*	Well, I like the Bond Girl. ·····
6	**A→R**	**2nd**	*Pre*	I see.
7	**R→A**	**1st**	*Pre*	I think that movie is good because of the setting of the "old age" for the 44-year old James Bond. ····· ❷
8	**A→R**	**2nd**	*H*	Uh-huh. ····· ❸
				(R is approved to obtain an initiative)
9	**R→A**	**3rd**	*H*	Yes. ····· ❹
10	**R→C**	**1st**	*H*	Have you ever seen the "Skyfall"? ❺
11	**C→R**	**2nd**	*H*	No, I haven't. ····· ❻
12	**A→C**	**1st**	*H*	Oh, you haven't seen it?
13	**C→A**	**2nd**	*H*	I never seen that before.

Figure 3-17: Interaction scenes. The "AP" signifies adjacency pair types. At **#4**, the system recognized A's adjacency third part and then generated a spontaneous opinion addressed to A (**#5**) as the first part. At that point, the system assumed the state of harmony (s_h) had changed from *Un-Harmonized* to *Pre-Harmonized*. After the system observed A's second part at **#8**, it assumed it at gotten approval to obtain an initiative to control the context (*Harmonized*). At **#10**, the robot asked C a question in order to give him the floor.

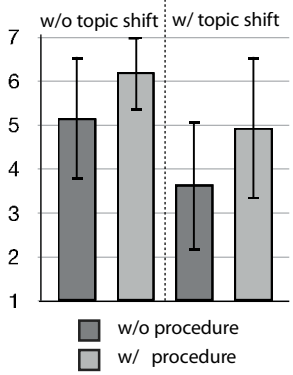


Figure 3-18: Result of experiment 1-a (appropriateness of procedures and topic shifting)

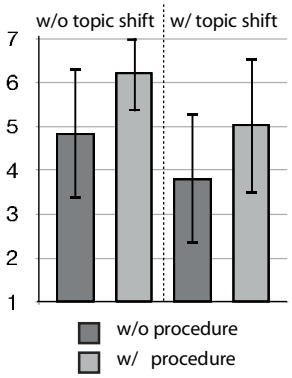


Figure 3-19: Result of experiment 1-b (groupness effects of procedures and topic shifting)

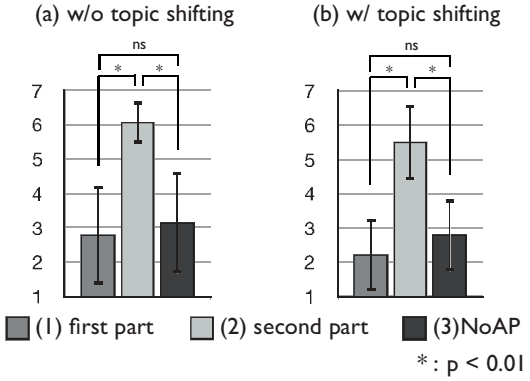


Figure 3-20: Result of experiment 2 (timing of initiating procedures)

Under conditions 1 and 2, the robot initiated its procedures just after the first and second parts, respectively. For condition 3, the robot initiated its procedure in the middle of the adjacency pairs, which is intended to show that the robot does not consider adjacency pairs. We did not consider the timing of the third part of the adjacency pair because we had already examined its appropriateness in experiment 1. After watching the videos, the participants were asked to answer 7-scale Likert questionnaires about the robot's **appropriateness of behavior**.

3.5.5 Results of Experiment 1 and 2

Figure 3-18 shows that the appropriateness of usage of procedures and topic shifting. The results of a two-way analysis of variance (ANOVA) show that there are significant differences among conditions in terms of both procedure ($F[1, 124] = 24.28$, $p < 0.01$) and topic shifting ($F[1, 124] = 34.19$,, $p < 0.01$). Figure 3-19 shows that the groupness effects of procedures and topic shifting. The results of a two-way analysis of variance (ANOVA) shows that there are significant differences among conditions in terms of both procedure

Table 3.13: Transition probabilities of adjacency pair parts used in experiment 3

From \ To	1st	2nd	3rd
1st	0.14	0.82	0.05
2nd	0.07	0.15	0.78
3rd	0.37	0.22	0.41

($F[1, 124] = 28.82$, $p < 0.01$) and topic shifting ($F[1, 124] = 21.09$, $p < 0.01$).

Figure 3-20 (a) shows that initiating procedures without topic shifting just after the second pair parts is more appropriate than other conditions. The results of an analysis of variance (ANOVA) show significant differences among conditions ($F[2, 26] = 34.46$, $p < 0.01$). The results of multiple comparisons using the Tukey HSD method show a significant difference between conditions 1 and 2, as well as between conditions 2 and 3 ($p < 0.01$). Figure 3-20 (b) shows that initiating procedures with topic shifting just after the second pair parts is more appropriate than the other conditions. The results of an ANOVA show significant differences among conditions ($F[2, 26] = 42.52, p < 0.01$). The results of multiple comparisons using the Tukey HSD method show a significant difference between conditions 1 and 2, as well as between conditions 2 and 3 ($p < 0.01$).

These results indicate that the usage of procedures to obtain initiative before approaching an un-harmonized participant showed evidence of acceptability and feeling of groupness. Regarding timing, initiating the procedures just after the second or third adjacency pair part is considered more appropriate than that after the first pairs.

3.5.6 Experiment 3: Evaluation of POMDP via User Simulation

This section describes comparison of POMDP and MDP-based group maintenance procedures using a user simulator. The purpose of this experiment was to evaluate how a robot could properly approach an un-harmonized participant under recognition error conditions using POMDP model. In order to focus on evaluating how a robot could reach an un-harmonized participant to get him/her harmonized without breaking conversational norms as soon as an un-harmonized situation is detected, we assumed that emergence of an un-harmonized participant could be detected with 100% accuracy (participation role recognition and motivation estimation modules) in this experiment. We only controlled the accuracy of adjacency pair recognition, which is the most critical factor to achieve a goal without breaking conversational norms.

The rewards were defined as Table 3.9. Each policy is trained by the HSVI algorithm described in Section 3.3. In this experiment, we assumed the system should obtain an initiative as soon as it got an opportunity to approach an un-harmonized (left behind) participant, and make him/her harmonized with a question. We constructed this experiment using the following user simulator. Each 10 dialogue act turns makes one unit, and each trial was performed with 1000 dialogue units (at most 10,000 dialogue act turns). An un-harmonized participant (always person C) emerged in the beginning of each turn. The user simulator returns one participant action A_p, as described in Table 3.7, on the basis of the previous system action. If the system could approach an un-harmonized participant with a question action (*question-current-topic* or *question-new-topic*) properly, the unit would be successful. We evaluated system performances with shifting observation probability (equation (3.10)), while the observation probability of the un-harmonized participant's motivation was always 1 (no errors) in order to generate an un-harmonized participant precisely in the beginning of each unit.

Table 3.14: An example sequence of the user simulation experiment using POMDP. Each row represents each turn. The "T/F" column represents whether "*question-current-topic*" was selected properly or not.

		Actual Situations and Ideal System Actions		Observations and Selected System Actions		
Turn	Motivation S'_m	Participant Action	System Action	Participant Action A'_p	System Action A'_s	T/F
1	C-Motivated	B-third	nod/null	*B-first* (error)	none	
2	C-Motivated	A-first	opinion	A-first	opinion	
3	C-Motivated	B-second-toR	qCur/qNew	B-second-toR	**qCur**	T

Table 3.15: An example sequence of the user simulation experiment using MDP.

		Actual Situations and Ideal System Actions		Observations and Selected System Actions		
Turn	Motivation S'_m	Participant Action	System Action	Participant Action A'_p	System Action A'_s	T/F
1	C-Motivated	B-third	nod/null	B-third	none	
2	C-Motivated	A-first	opinion	A-first	opinion	
3	C-Motivated	A-first-toR	answer	*A-second-toR* (error)	**qCur**	F

Table 3.14 and 3.15 shows example sequences of the user simulation experiment using POMDP and MDP, respectively. In **#1** of Table 3.14 (a beginning of an unit), an un-harmonized person was observed. In this turn, while a person B's first pair part (an error sensory input) were observed (A'_p = "B-first"), the system did not select a dialogue action, but just looked at person B (A'_s = "none") according to a lower confidence score. In **#2**, the system generated its own opinion along a current topic (A'_s = "opinion"), just after a observation of person A's first pair part (A'_p = "A-first"), to initiate a procedure according to a higher confidence score. Then, in **#3**, person B's second pair part reacting to the system's previous action (A'_p = "B-second-toR") was observed, and the system finally generate a question to the un-harmonized person (person C) along a current topic (A'_p = "qCur") to give him a turn properly. Now this unit was successful ("T" in the "T/F" column represents "succeeded").

While the POMDP-based system could cope with errors, the MDP-based system was more sensitive to observations, therefore acts aggressively. In **#1** of Table 3.15 (MDP), an un-harmonized person was observed. In **#3**, while person A actually asked the system with a question action regarding the system's previous opinion (A'_p = "A-first-toR"), the system observed "A-second-ToR" (an error sensory input). Then, the system naively selected a question action to person C (A'_s = "qCur"), ignoring A's question, and eventually, this unit failed (there were no opportunities afterward to selecting "qCur" action again within this unit).

We evaluated each unit whether a system could select a question action to an un-harmonized participant, with comparing ideal system actions. We calculated precision and recall of each question action (*question-current-topic* or *question-new-topic*) as follows:

$$Precision = \frac{|\{Correctly\_Question\_Selected\} \wedge \{Actually\_Question\_Selected\}|}{|\{Actually\_Question\_Selected\}|} \tag{3.17}$$

$$Recall = \frac{|\{Correctly\_Question\_Selected\} \wedge \{Ideally\_Question\_Should\_Be\_Selected\}|}{|\{Ideally\_Question\_Should\_Be\_Selected\}|} \tag{3.18}$$

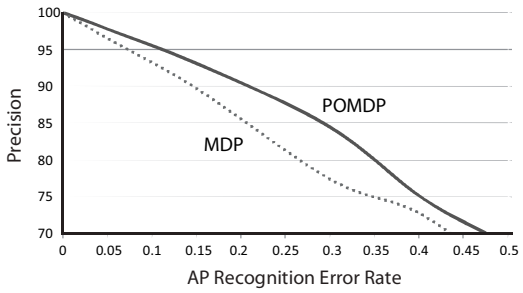


Figure 3-21: Precision of timing of initialing a procedure

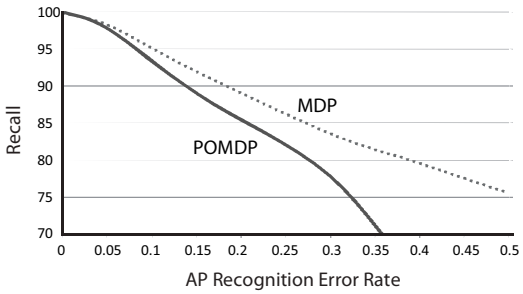


Figure 3-22: Recall of timing of initialing a procedure

The precision and recall as an adjacency pair observation error rate for the two types of systems are shown in Figure 3-21 and Figure 3-22, respectively. The general trend is that precision of the POMDP is better than that of the MDP-based procedure; however, this is the opposite for recall.

3.6 Conclusions and Future Work

3.6.1 Summary and Contributions

We proposed a framework for conversational robots harmonizing four-participant groups. Based on a representation of conversational situations, we presented a model of procedures obtaining conversational initiatives in incremental steps to harmonize such four-participant conversations. These situations and procedures were modeled and optimized as a partially observable Markov decision process (POMDP). As the results of two user experiments, usages of procedures obtaining initiatives showed evidences of acceptability as a participant's behaviors, and feeling of groupness. As for timings, initiating the procedures just after the second or third adjacency pair parts is felt more appropriate than the first pairs by participants. And the result of the simulation experiment, POMDP showed reasonably better performance for group maintenance than MDP. Because of the robustness of POMDP, it's suitable for procedural group maintenance, including its timing to begin a procedure. The main contribution of this research is that we modeled a facilitation model in 4-participant conversational situations, which is the minimum unit of facilitation process. We indicated and defined "harmony of conversation"' based on "engagement density" and status of interest sharing.

3.6.2 Extensions of POMDP

The future work include considering extensions of POMDP model for task goal management, while we discussed mainly aspects of group maintenance for facilitation in this paper. Williams et al. presented the POMDP-based spoken dialogue system (SDS-POMDP), where they modeled the user goal of a task. Based on the idea, we will consider the goal model of a group task for optimization considering longer term rewards. Also, in order to deal with situations of more than four participants, some approximation methods for larger state space of POMDP should be considered.

3.6.3 Extensions of Situation Understanding

We are also considering extending the modules of the situation understanding process, including participation role recognition and motivation estimation based on advantages of related work.

Many research mentioned that acoustic and visual cues, such as gaze direction, face direction, head pose and acoustic information are reliable cues for addressing in multiparty human-human and human-robot interactions (Katzenmaier et al., 2004). Jovanovic et al. presented results for addressee identification in four-participant face-to-face meetings (Augmented Multiparty Interaction (AMI) meeting corpus (Carletta et al., 2006)). Their classifiers performed best with a combination of conversational context including adjacency pairs, and utterance features and speaker gaze information, using a Bayesian Network and Naive Bayes classifiers (Jovanović et al., 2004; Jovanovic et al., 2006; Jovanović et al., 2006). Based on the Kendon's finding that speakers look away at beginning of turns, and look back at their interlocutors towards end of turns (Kendon, 1967), Fujie et al. proposed gaze recognition for turn-taking model with a conversational robot (Fujie et al., 2006), and Johansson et al. also showed that the current speaker only look at the next speaker at the end of the turn in one fourth of the cases if it was a human and almost half of the time if it was the robot (Johansson et al., 2013). In our current study, we assumed only one addressee at the time as simplification. Extension of addressing model allowing for two or more addressees remains as an open question.

As for the motivation estimation module, there are many works on modeling participant's internal states including interests and emotions, relevant to concepts to our motivation model. Gatica-Perez et al. presented an investigation of the performance of audio-visual fusion cues on classifying high v.s. neutral group interest-level segments using HMM based methods (Gatica-Perez et al., 2005). Because we believe these internal model of participants would be necessary components for model of multiparty conversation, evaluations of our motivation module in real conversational situations also will be our future work.

CORRESPONDENT: "Which of the cities visited did Your Highness enjoy the most ?"

PRINCESS ANN (Audrey Hepburn): "Each, in its own way, was unforgettable. It would be difficult to — Rome! By all means, Rome. I will cherish my visit here in memory as long as I live."

"Roman Holiday"

4 Language Generation

We present the SCHEMA QA, an enjoyable question answering framework that has expressive opinion generation mechanisms. In terms of functional conversations, Grice's Maxim of Quantity suggests that responses should contain no more information than was explicitly asked for. However, in our daily conversations, more informative response skills are usually employed in order to hold enjoyable conversations with interlocutors. These responses are usually produced as forms of one's additional opinions, which usually contain their original viewpoints as well as novel means of expression, rather than simple and common responses characteristic of the general public. In this paper, we propose automatic expressive opinion sentence generation mechanisms for enjoyable conversational systems. The generated opinions are extracted from a large number of reviews on the web, and ranked in terms of contextual relevance, length of sentences, and amount of information represented by the frequency of adjectives. The sentence generator also has an additional phrasing skill. Three controlled lab experiments were conducted, where subjects were requested to read generated sentences and watch videos filmed about conversations between the robot and a person. The results implied that mechanisms effectively promote users' enjoyment and interests.

4.1 Introduction

We present the SCHEMA ([$\int$ *e:ma*]) QA, an enjoyable question answering framework comprising informative sentence generation mechanisms. Let us begin by looking at an example of a question asked by one person and answered by another:

A: "*Do you have any favorite actor or actress?*"
B: "*Yes, my favorite actress is Audrey Hepburn. Audrey is, just as one would expect, a charming and beautiful woman even in her private life!*"

In this example, person A asked about person B's favorite actor/actress; person B answered the question in the first sentence, and continued by adding another opinion about Audrey Hepburn along the current context, an utterance which person A may not have expected. In terms of functional conversations, Grice (Grice, 1975) described the cooperative conversation principle as consisting of four maxims (Quality, Quantity, Relevance, and Manner) that arise from the pragmatics of natural language. Grice's Maxim of Quantity suggests that responses should contain no more information than was explicitly asked for. Seen from this viewpoint, person B's additional opinion above contradicts the maxim because it resulted in too much information being given. However, in our daily conversations, more informative phrasing and response skills are usually employed in order to hold enjoyable conversations with interlocutors. Below, we analyze elements of enjoyable conversations at both the discourse and the sentence level.

At the discourse level, structures of enjoyable conversations including additional own opinions to keep the thread of the conversation, just like the example above, are associated with "small talk" skills. The phenomenon of small talk was initially studied by Malinowski (Malinowski, 1994), who coined the term "phatic communication" to describe "a type of speech in which the ties of union are created by a mere exchange of words," which is a mechanism for managing the engagement of communication and psychological distance among interlocutors. Small talk is generally used as a conversation opener, at the end of a conversation, and as a space filler to avoid silence. Through the utility of small talk, we not only accomplish specific tasks but also enjoy the conversations themselves. Schneider (Schneider, 1988) did the first extensive study of small talk. He theorizes that such a conversation consists of a number of "moves": topic initialization, agreeable phrasing, informative responding, and acknowledgement. According to Schneider's categorization, person B's action of additionally responding with his/her own opinion in the example above can be regarded as an informative responding move.

At the sentence level, person B's simple opinion, "*Audrey is beautiful, isn't she?*," for example, may have less information and sufficiently meet the requirement of Grice's Maxim. This is not only due to the length of the sentence, but also because it is a common opinion in line with that of the general public. Therefore, it cannot attract an interlocutor's interest in an effective manner. In contrast, person B's actual second sentence above, "*Audrey is, just as one would expect, a charming and beautiful woman even in her private life*," expresses a novel opinion about Audrey Hepburn with a wealth of words and from an original viewpoint. It also implicitly contains its reason for the (positive) attitude. Let us examine a few more examples:

> "*This is the **erotic** thriller movie which also expresses the elegance of the ballet.*"
> "*Dola's family and Princess Sheata with **pure** mind are really **cute**, **charming** and, **innocent**.*"

These two opinions are about the movies "*Black Swan*" and "*Castle in the Sky*," respectively. Adjectives are shown in boldface. Sentences with less frequently used adjectives and proper length are likely to provide a unique viewpoint that is different from that of the majority, in contrast to redundantly long sentences, which may cause harmful effects instead. From this analysis, we assume that the amount of information in each sentence can be described as a level of generality of expressions, which is mostly represented by the frequency of adjectives in documents on a certain topic, the number of adjectives in each sentence, noun relevance to the current context, and the length of each sentence. Some natural language processing techniques, namely, sentiment analysis(Turney, 2002; Pang et al., 2002) and opinion mining (Nakagawa et al., 2008), have a direct relation to opinion generation. Their motivations are mostly to analyze users' preferences automatically extracted from a large amount of review data for marketing and service improvements. However, there are very few works that apply opinion generation to dialogue systems in terms of novelty of the sentences themselves.

On the basis of the results of these analyses, we propose an enjoyable question answering framework that is capable of small talk, including additional phrasing skills and automatic expressive opinion generation. The opinions are extracted from a large number of reviews on the web, and ranked in terms of contextual relevance, length of sentences, and amount of information represented by the frequency of adjectives. The additional phrasing skill is implemented as a mechanism of sentence combination of a simple preceding response and an additional opinion. Our typical scenario is as follows: When the system is asked a factoid-typed question, it first replies with a sufficient answer based on a structured database, and then it adds another expressive opinion in line with the current context, which might be informative to the user.

The remainder of this paper is organized as follows: In Section 4.2, we review related work done on sentence generation and small talk skills. In Section 4.3, we describe the automatic sentence generation process, inclusive of opinion extraction and ranking, and give an overview of our proposed question answering framework in Section 4.4. In Section 6.5, we discuss the results of three experiments conducted to determine the effectiveness of enjoyment of sentence and additional phrasing. Finally, in Section 6.6, we conclude this paper and outline future research directions.

4.2 Theoretical Framework of Language Generation for Enjoyment

In this section, we review works done in relation to the production of enjoyable conversations at both the discourse and sentence levels. At the discourse level, we outline studies dealing with small talk, which have mostly been discussed in linguistics and socio-psychology. We also review conversational agent systems along with small talk skills. At the sentence level, we give an overview of opinion mining and sentiment analysis, which are major tasks in natural language processing.

4.2.1 Small Talk

The phenomenon of small talk was initially studied by Malinowski as "phatic communication" (Malinowski, 1994). It usually occurs as *conversation opener*, *the end of a conversation*, *space filler to avoid silence* (Holmes, 2000). Throughout a small talk, we not only achieve specific tasks but also know his/her interlocutor and enjoy conversations for themselves. Schneider carried out the first extensive study of small talk (Schneider, 1988). He theorized that such a conversation consists of a number of fairly predictable segments or "moves," which can be summarized as follows: *topic initialization*, *agreeable phrasing*, *acknowledgement*. Based on these early studies, Bickmore and Cassel (Bickmore and Cassell, 1999) pioneered an embodied conversational agent that uses small talk skills to engage a rapport with interlocutors in the domain of real estate sales (REA). They also describe an exercise adviser agent that promotes exercise among students in order to evaluate the effectiveness of relational agents in health behavior change. Consequently, they found significantly more laughter in conversations in which small talk occurred compared to those dialogs from which small talk was virtually absent (Bickmore, 2003). In their system, interaction begins in the greet frame, then automatically transitions into the small talk frame. Within the small talk frame, the dialog is conducted with segments and moves modeled according to Schneider's moves (specifically, *Topic initialization*, *Agreeable phrasing*, and *Acknowledgment*). An example interaction in their REA system is as follows (comments in parentheses represent our interpretation of small talk move):

——————— Small Talk Frame ———————

`System`: "*How about this weather?*" (Topic initialization)

`User`: "*It's gorgeous.*"''

`System`: "*It s gorgeous?*" (Acknowledgement)

`User`: "*Don't you think it's a wonderful Spring?*"‘’
`System`: "*Maybe.*"
`User`: "*Yea?*"
`System`: "*Yea.*" (Acknowledgement)
———————— End of Small Talk Frame ————————

Dosaka et al. (Dohsaka et al., 2009) examined the method of conversational agents' acknowledgment. They developed a thought-evoking dialogue system that can express acknowledgment of a user's dialogue acts with empathic expressions, which was implemented as a quiz game task in multiparty conversation. They reported that the agents' empathic expressions are effective for users' satisfaction and an increasing number of user utterances. Our preliminary experiments conducted using a conversational robot examined the effectiveness of the robot's informative phrasing. The results from the experiments indicated that the combination of passive response and spontaneous phrasing was sufficiently effective to facilitate entertaining conversations (Matsuyama et al., 2011).

Topic selection usually depends on contexts incluting relationship between the two people and the environent of the conversation. The social penetration theory (Taylor and Altman, 1987) describes the ways relationship deepens, the breadth and depth of the topics disclosed become wider and deeper, helping the interlocutor to gain common ground. In early stages of relationship, "safe" topics such as *the weather*, *recent shared experiences*, *movies*, *foods* are preferrable to be initialized (Holmes, 2000).

Given this general small talk framework at the discourse level, we further discuss informative productions at the sentence level, specifically, subjective expression generation which have not been substantially discussed in previous small talk system research, in the next section.

4.2.2 Natural Language Generation Pipeline

Traditionally, natural language generation (NLG) systems consist of three major processes which are connected together in a pipeline: *content planning*, *microplanning*, *realization* (Reiter et al., 2000).

Content Planning

Content planning consists of content determination and document structuring. *Content determination* decides what information will appear in the output text. This depends on what your goal is, who the audience is, what sort of input information is available to you in the first place and other constraints such as allowed text length. *Document structuring* decides how chunks of content should be grouped in a document, how to relate these groups to each other and in what order they should appear. For instance, when describing last month ' s weather, you might talk first about temperature, then rainfall. Or you might start off generally talking about the weather and then provide specific weather events that occurred during the month.

Fabbrizio et al. proposed a content planning for review of restaurants using summarization techniques (Fabbrizio et al., 2013a,b, 2014). Higashinaka et al. proposed an unsupervised method for learning a dictionary mappings between the semantic representations of concepts and content plans from user reviews of restaurant and hotel domains (Higashinaka et al., 2007).

Microplanning

Misroplanning consists of both sentence and nonverbal language planning. Sentence planning consists of syntactic template selection, lexical selection, referring expressions generation and aggregation. Nonverbal

language planning consists of eye gaze planning, body orientation planning and iconic gesture planning. *Syntactic Template Selection* Particular syntactic structures are chosen as well. *Lexical Selection* decides what specific words should be used to express the content. For example, the actual nouns, verbs, adjectives and adverbs to appear in the text are chosen from a lexicon. *Referring expressions generation* decides which expressions should be used to refer to entities (both concrete and abstract). The same entity can be referred to in many ways. For example March of last year can be referred to as: "April 2014", "April of the previous year", "it". *Aggregation* decides how the structures created by document planning should be mapped onto linguistic structures such as sentences and paragraphs. For instance, two ideas can be expressed in two sentences or in one: "The month was cooler than average.The month was drier than average." v.s. "The month was cooler and drier than average."

There are several approaches for microplanning. Stone et al. proposed the SPUD (Sentence Planner Using Descriptions) that has a tree search algorithm for simultaneously constructing both the syntax and semantics of a sentence using a Lexicalized Tree Adjoining Grammar (LTAG). This approach captures naturally and elegantly the interaction between pragmatic and syntactic constraints on descriptions in a sentence, and the inferential interactions between multiple descriptions in a sentence. At the same time, it exploits linguistically motivated, declarative specifications of the discourse functions of syntactic constructions to make contextually appropriate syntactic choices[1] (Stone, 2002; Stone et al., 2003). Based on the SPUD, Cassell et al. proposed the generation of verbal and nonverbal communicative actions in an implemented embodied conversational agent. Their agent plans each utterance so that multiple communicative goals may be realized opportunistically by a composite action including not only speech but also nonverbal gesture that fits the context and the ongoing speech in ways representative of natural human conversation. They accomplished this by reasoning from a grammar describing gesture declaratively in terms of its discourse function, semantics and synchrony with speech (Cassell et al., 2000; Kopp et al., 2004).

Stent et al. proposed SPaRKy (Sentence Planning with Rhetorical Knowledge)[2] a sentence planner that uses rhetorical relations and adapts to the user's individual sentence planning preferences (Stent et al., 2004). SPaRKy receive a discourse plan as a input (a tree with rhetorical relations on the internal nodes and a proposition representing a text span on each leaf), and outputs one or more sentence plans (each a tree with discourse cues and/or punctuation on the internal nodes). SPaRKy employs "over-generate and select" way that has a two-stage sentence planning. In the the first stage, possible sentence plans are generated through a decisions process using only local information about single nodes in the discourse plan. In the second stage, the generated sentence plan candidates are ranked using a user/domain specific sentence plan ranker, which evaluates the global quality of each sentence plan (Walker et al., 2007). Sentence plan generation proress in SPaRKy consists of four tasks: span ordering, sentence aggregation, and discourse cue selection, and a simple referring expression generation. They also automatically extracted planning rules from RST-DT corpus (Stent and Molina, 2009).

Mairesse et al. proposed the PERSONAGE (PERSONAlity GEnerator)[3], a highly parametrizable sentence generator to change personalities of an agent. They applied extraverted and intraverted personalitied, based on the "Big Five" personality model. The planner has many parameters in the stages of syntactic templates selection, aggregation Operations, pragmatic transformation and lexical choice (Mairesse and Walker, 2007).

[1] http://www.cs.rutgers.edu/ mdstone/class/taglet/

[2] http://www.research.att.com/archive/people/Stent_Amanda_J/library/documents/sparky2.0/index.html

[3] https://games.soe.ucsc.edu/project/personage

Realization

Realization consists of linguistic realization and structure realization. *Linguistic realization* uses rules of grammar (about morphology and syntax) to convert abstract representations of sentences into actual text. *Structure realization* converts abstract structures such as paragraphs and sentences into mark-up symbols which are used to display the text. RealPro(Lavoie and Rambow, 1997) and SimpleNLG (Gatt and Reiter, 2009) perform the realization process.

4.2.3 Opinion Mining and Sentiment Analysis

Although small talk is enjoyed for itself, as discussed above, there are few considerations on expressive sentence generation for enjoyment. In this research, in contrast to the general pipeline of natural language generation, we attempt to apply information retrieval approach to the purpose. One of the related methods of automatic opinion generation is opinion mining and sentiment analysis, one of the major applications of natural language processing to identify and extract subjective information from source materials. In general, opinion mining and sentiment analysis aims to identify a speaker or a writer's subjective attitude with respect to a certain topic or a contextual polarity of a document. The recent rise of social media has fueled interest in sentiment analysis. A basic task in sentiment analysis is to classify the polarity (e.g., positive, negative, or neutral) of a given text at the document, sentence, or feature/aspect level. With their early works in this area, Turney (Turney, 2002) and Pang (Pang et al., 2002) applied different methods to documents to detect the polarities of product reviews and movie reviews, respectively. Turney presented a simple unsupervised learning algorithm for classifying reviews, while Pang classified a document's polarity on a multi-way scale.

At the sentence level, the major tasks are building evaluative dictionaries, evaluative expressions extraction, and subjectivity/objectivity identification. Evaluative dictionaries aim at building sets of expressive words and emotional polarity. Each generated dictionary has a wide range of applications, including predicting the emotional polarity of sentences and documents. Kamps et al. (Kamps et al., 2004; Fellbaum, 2010) developed a distance measure for the semantic orientation of adjectives by investigating a graph-theoretic model of WordNet's synonymy. Nasukawa et al. (Nasukawa and Yi, 2003) used context information around the subject term. Kobayashi et al. (Kobayashi et al., 2005) structured dictionaries with 5,500 entries from reviews (230,00 sentences in total) using the semi-automatic method and additional expansion by hand. Higashiyama et al. (Higashiyama et al., 2008) structured a Japanese evaluative noun dictionary using selectional preferences. Evaluative expressions extraction is based on these evaluative dictionaries. In order to extract evaluative expressions that can appear at any position in a sentence, Nakagawa et al. use the BIO encoding method, which has been commonly used for extent-identification tasks (Breck et al., 2007; Sha and Pereira, 2003) Subjectivity/objectivity identification is defined as classification problem where a given text is classified into objective or subjective classes. (Pang and Lee, 2008). This problem is sometimes difficult because the subjectivity of words and phrases may depend on their context, also objective and subjective sentences may be intermingled in a text. In this paper, we focus on extracting the opinions themselves in Japanese, even though the goals of the methods above are for sentiment polarity analysis, not to extract the opinions themselves. We utilize Nakagawa's extraction methods and the expressive dictionaries proposed by Kobayashi and Higashiyama. In addition, we regard all opinion candidates extracted from a review site as subjective opinions.

In terms of novelty and serendipity of a system's production, some preliminary discussions exist in the recommender systems research domain. Herlocker et al. presented various metrics, including novelty and serendipity beyond recommendation accuracy, to evaluate users' satisfaction with recommender systems

(Herlocker et al., 2004; McNee et al., 2006a,b). Noda et al. proposed a general method for extracting serendipitous information from Wikipedia that uses its network structure (Noda et al., 2010). In this paper, we discuss elements of novelty and serendipity of one opinion. As we discussed in Section 4.1, novelty of expressions and serendipity of viewpoints are mostly represented by the frequency of adjectives, which has not been substantially considered.

In the following sections, we first present our proposed method for automatic expressive opinion generation in Section 4.3, and then present a system architecture that enables small talk skills utilizing generated opinions in Section 4.4.

4.3 Expressive Opinion Generation

Our proposed opinion sentence generation system consists of four processes: document collection, opinion extraction, sentence style conversion, and sentence ranking.

4.3.1 Document Collection

Topics in small talk are considered to be "safe" in most circumstances (Holmes, 2000). They include *the weather*, *recent shared experiences*, *movies*, *foods*, and so on. In this paper, we employ topics from the *movies* domain. We collected review documents from the Yahoo! Movie site [4] with a review crawler we implemented. These reviews are preliminarily sorted by users' ratings (five-star rating system) because those reviews with the higher ratings are more likely to contain positive opinions. We regarded the top one thousand reviews as our target documents. Table 4.1 displays an example of review sentences about the movie "*Castle in the Sky*." We decided to select these sites because of both substantial volume and quality of reviews.

4.3.2 Opinion Extraction

Opinion extraction comprises two processes: extraction of evaluative expressions and classification of their sentiment polarities (positive/negative). We eliminate opinions with negative sentiments because a system is expected to talk about positive contents in our conversational task. Particular words such as "like" and "hate" are often used to express evaluation, which are related with certain sentiment polarities. We use both a subjective evaluative dictionary (Kobayashi et al., 2005) and an evaluative noun dictionary (Higashiyama et al., 2008)[5]. The subjective evaluative dictionary contains words such as "comfortable" and "regrettable" with their sentiment polarities. The evaluative noun dictionary contains nouns with desirable and undesirable properties, such as "health" and "cancer," respectively. We also use our additional hand-crafted dictionary using corpus we collected from reviews on the Yahoo! Movie site. On the basis of the method proposed by Nakagawa et al. (Nakagawa et al., 2008), we use linear-chain conditional random fields (CRF) for the BIO encoding. Using BIO, each word is tagged as either (B)eginning an entity, being (I)n an entity, or being (O)utside of an entity. We use the CRF++ toolkit[6] in our experiments. An example of BIO encoding is shown in Figure 4-1. The sentence means "I watched the movie *Roman Holiday* the other day. Audrey is beautiful, isn't she?" In this case, the segment "Audrey is beautiful, isn't she?" is an evaluative expression. The following features are used to predict the BIO tags of the i-th word in a sentence:

[4] http://movies.yahoo.co.jp/
[5] http://www.cl.ecei.tohoku.ac.jp/
[6] https://code.google.com/p/crfpp/

Table 4.1: Example of reviews for "*Castle in the Sky*."

[Reviewer 1]: 必見. (Must see.)
[Reviewer 2]: 子供の頃に見て，シータのあの青い石のペンダントが欲しかったな. (I wanted the blue stone pendant when I watched it as a child.) キラキラ青く輝いて，宙に浮いていたし. (It was brilliantly scintillating and up in the air.) 何回観ても面白いです. (No matter how many times I watch it, it is still fun.)
[Reviewer 3]: 宝探しという冒険を現代の子供たちは思い描く事が出来るのだろうか. (I am wondering if kids today can imagine what treasure-hunting adventures are like.) 高層ビルとアスファルトに覆われた現代社会に生活する事が悪いとは言わないが少なくとも子供のピュアな精神を培う上に於いては良いとは言えないだろう. (I am not saying that living in a modern society is bad, but it is not good for nourishing children's pure hearts.) 本作品はそんな現代人の精神の欠乏した部分を見事に埋め合わせ修復してくれる作品である. (This movie can beautifully restore the spirit currently lacking in contemporary society.) 子供が見れば尚の事，類い稀なる精神を養う事に繋がると思う. (It helps children's spirits to be extraordinary.)
[Reviewer 4]: ナウシカと趣がちがうものの，やはり最高の映画です. (Although it is different in flavor from "*Nausicaa*," it's still a masterpiece.)
[Reviewer 5]: まさに冒険活劇の王道をいくストーリー. (Its story takes the high road of adventure action pictures.) 理屈抜きでのめり込める. (It viscerally makes us go deep into the story.) ヘタに道徳的思想なんか織り込もうとしないで，宮崎監督にはこういうストレートで単純な作品を作ってもらいたい. (I want Mr. Miyazaki to make this kind of straightforward and pure movie, and not embed poor morals in them.)

$$
\begin{aligned}
&s_{i-2}, s_{i-1}, s_i, s_{i+1}, s_{i+2}, s_{i-1}\&s_i, s_i\&s_{i+1},\\
&b_{i-2}, b_{i-1}, b_i, b_{i+1}, b_{i+2}, b_{i-1}\&b_i, b_i\&b_{i+1},\\
&c_{i-2}, c_{i-1}, c_i, c_{i+1}, c_{i+2}, c_{i-1}\&c_i, c_i\&c_{i+1},\\
&f_{i-2}, f_{i-1}, f_i, f_{i+1}, f_{i+2}, f_{i-1}\&f_i, f_i\&f_{i+1},\\
&p_{i-2}, p_{i-1}, p_i, p_{i+1}, p_{i+2}, p_{i-1}\&p_i, p_i\&p_{i+1}
\end{aligned}
$$

where s_i, b_i, c_i, f_i, and p_i denote the surface form, the base form, the coarse-grained part-of-speech (POS) tag, the fine-grained POS tag, and the polarity, to the i-th wording of the input sentence, respectively. "&" symbol indicates a conjunction features. We used our hand-crafted opinions and typical segment of opinions collected from review sites as a corpus for learning.

Judgment of the evaluation polarity is a method used to detect the bias of the evaluation sentence and can reject negative opinion sentences. In this paper, we refer to the method from Nakagawa's study (Nakagawa et al., 2010). This method is a dependency tree-based method for sentiment classification of subjective sentences using conditional random fields with hidden variables. For example, in Fig.4-2, "cancer" and "heart disease" have themselves negative polarities. However, the syntactic dependency of "prevents" inverts the

polarity and the sentence is classified as positive polarity. In the figure, each phrase in the subjective sentence has a random variable. The random variable represents the polarity of the dependency subtree whose root node is the corresponding phrase. The node denoted as <*root*> indicates a virtual phrase which represents the root node of the sentence, which is regarded that the random variable of the root node is the polarity of the whole sentence. Nakagawa et al. reported the precision, the recall (positive polarity), and F-value of whole sentence polarity were 0.87, 0.79, and 0.89, respectively. Table 4.2 displays our examples of extracted evaluation sentences from "Castle in the Sky." In this table, "pol" represents the polarity of a whole sentence. All the sentences classified as negative will be eliminated in this process in order to maintain high precision, even if it would decrease recall, to give the highest priority to safety. We used JUMAN[7], as a Japanese language morphological analyzer.

4.3.3 Sentence Style Conversation

In order to preserve the consistency of the system's character, we convert the style of the sentences. We focus on expressions at the end of Japanese sentences, such as question tags and formal/casual lines, because character styles primarily appear in this part of Japanese sentences. In our experimental system, we convert them into casual and empathic styles. For example, the last part of an original formal sentence such as

"良いと思います (*Yoi to omoi masu*)"

can be converted into

"良いと思うんだ (*Yoi to omou nda*)",

which sounds more casual in Japanese. However, both mean "I think it's good."

The sentence style conversion process is based on a handwritten rule we prepare. After Japanese morphological analysis, punctuation marks and special symbols are eliminated. The last morpheme is converted based on part of speech. For example, a part of the sentence "良いと思います (*Yoi to omoi masu*)" can be analyzed as

"良い (adjective) / と (particle) / 思い (verb) / ます (postfix)".

In this case, the verb "思い (*omoi*)" can be stemmed as "思う (*omou*)," and a new postfix "んだ (*nda*)" appended. This results in "良いと思うんだ (*Yoi to omou nda*)." Examples of converted sentences from the topic "*Castle in the Sky*" are shown in Table 4.3.

4.3.4 Sentence Ranking

In this section, the scales used to rank the sentences are explained, and ranking methods are introduced. The scale consists of three components, such as the importance of the word, adjective frequency, and number of morphemes. The importance of the word is a degree of the relation between the sentence and the topics. adjective frequency is the scale for unexpectedness. The number of morphemes is the scale for extracting short clear sentences and long well-grounded sentences.

この	/	間	/	ローマ	/	の	/	休日	/	を	/	観た	/	んだ	/	けど	/	、	/	オードリー	/	が	/	綺麗だ	/	よ	/	ね	/	。
O		O		O		O		O		O		O		O		O		O		B		I		I		I		O		O

Figure 4-1: An example of BIO encoding. The sentence means "I watched the movie *Roman Holiday* the other day. Audrey is beautiful, isn't she?"

[7] http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN

Next, adjective frequency, a scale for unexpectedness for users, is discussed below. The frequency of adjectives affects the unexpectedness and expectedness. Utterances with high frequent adjective terms are expected to be common for users because many reviewers express themselves in the same way. For example, high frequent adjective terms in the topic "Audrey Hepburn" is "beautiful," which is simply a typical expression used for her. On the other hand, utterances with low frequent adjective terms are rare and the expressing of these inform the novelty. For example, the term "spirited" is a low frequent adjective term and so this expression is unexpected. Hence, sentences with low frequent adjective terms are given the attribute "unexpected."

We propose three rankings for algorithms in terms of length and novelty: **Short**, **Standard**, and **Diverse**. The ranking process is shown in Figure 4-3. As for the length of sentences, based on our experiences, we assume a sentence consisting of from seven to ten morphemes expresses the opinion clearly, and a sentence consisting of from fifteen to twenty are possibly expressing the opinion. A **Short** algorithm delivers an opinion to users elliptically with a short sentence. For example, in the topic "*Roman Holiday*," a **Short** sentence can be one such as "Audrey Hepburn had a gorgeous presence." In a **Short** algorithm, at first, we filter sentences in terms of the topic relativeness and the number of morphemes. As the topic relativeness, in this paper, we employ the top 30% of sentences sorted in terms of the term frequency-inverse document frequency (TF-IDF) scores. TF-IDF is the product of two statistics, term frequency and inverse document frequency. It is calculated as follows:

$$TF(w,d) = \frac{C(w,d)}{\sum_w C(w,d)} \tag{4.1}$$

$$IDF(w) = \frac{|D|}{|\{d \in D : w \in d\}|} \tag{4.2}$$

$$TF - IDF(w) = TF(w,d) \cdot IDF(w) \tag{4.3}$$

where $C(w,d)$ represents the frequency of term w in a document d, $|D|$ represents the total number of documents in the corpus, $|\{d \in D : w \in d\}|$ represents the number of documents in which the term w appears at least once ($TF(w,d) \neq 0$). We collected the top 64 web sites as 64 separate documents for each topic word using Google web search. We assume the top 30% of candidates are reasonably related with the current topic, according to a result of a preliminary experiment described in Section 4.5.3. So, the topic related sentences consisting of seven to ten morphemes are extracted. If a sentence has more than two nouns, one with the largest TF-IDF is employed as a representative.

Next, the top 30% list is sorted by adjective frequency. At this point, sentences that have only one adjective are employed because we assume more than two adjectives are too redundant for the **Short** sentences to state opinions briefly.

We assume sentences extracted by the **Standard** algorithm contains substantial opinions or reasons,

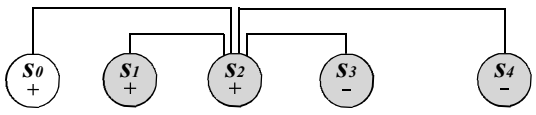


Figure 4-2: Probabilistic model based on dependency tree (Nakagawa et al., 2010)

Table 4.2: Extracted opinions and sentiments from "*Castle in the Sky*" after the polarity classification process.

Sentence	Pol.
子供の頃に見て，シータのあの青い石のペンダントが欲しかったな. (I wanted the blue stone pendant when I watched it as a child.)	+
何回観ても面白いです. (No matter how many times I watch it, it is still fun.)	+
宝探しという冒険を現代の子供たちは思い描く事が出来るのだろうか. (I am wondering if kids today can imagine what treasure-hunting adventures are like.)	+
本作品はそんな現代人の精神の欠乏した部分を見事に埋め合わせ修復してくれる作品である. (This movie can beautifully restore the spirit currently lacking in contemporary society.)	+
子供が見れば尚の事，類い稀なる精神を養う事に繋がると思う. (It helps children's spirits to be extraordinary.)	+
ナウシカと趣がちがうものの，やはり最高の映画です. (Although it is different in flavor from "*Nausicaa*," it's still a masterpiece.)	+
理屈抜きでのめり込める. (It viscerally makes us go deep into the story.)	+
ヘタに道徳的思想なんか織り込もうとしないで，宮崎監督にはこういうストレートで単純な作品を作ってもらいたい. (I want Mr. Miyazaki to make this kind of straightforward and pure movie, and not embed poor morals in them.)	-
宮崎作品で，傑作の一つにあげられると断言します. (I can definitely assert that this movie is one of the most awesome movies from Ghibli.)	+
ストーリー，キャラクター，演出その他全てに置いて正に完璧な作品. (This movie is just perfect in story, character, direction, and everything else.)	+
これを見た事が無いあるいは嫌いという奴は，日本人と認めない. (I don't believe there is anyone in Japan who has never seen this movie or does not like it.)	-
パズーとシータは永遠の少年と少女です. (Pazu and Princess Sheeta are "the" boy and "the" girl.)	+
夢冒険いつまでも忘れない作品です. (This movie always reminds us to dream of adventures.)	+
当時はまだ子供であまり理解していなくて，ただ激しいアニメーションに楽しみながら見ていたのを覚えている. (I just remember watching this movie with the dramatic animated actions without deep understanding when I was young.)	-
この作品は見れば見るほどその素晴らしさが感じられる. (The more you watch this movie, the better you get the feeling of its excellence.)	+
展開もいいのだが，１つ１つのシーンに無駄がなく感じられる. (I think there are no wasted scenes, and the plot is good.)	-
振り返ってみてみると，本当に１つ１つのシーンに意味があるし，１つ１つのシーンに思い出が詰まっている. (Each scene has meaning and I remember all the scenes.)	+
最初，シータが船から落ちるシーンや，町での喧嘩のシーン，ムスカにシータを奪われて，町に帰るまでのシーン. (The scene in which Princess Sheeta falls from the ship, the scene of fighting in the town, the scene in which Princess Sheeta is kidnapped by Colonel Muska and Pazu returns to the town.)	+
１つ１つが全て無駄なくつながっている感じがする. (All of those scenes are linked together with no waste.)	-
テンポがいい展開. (The tempo at which the plot of the story develops is good.)	+

which can appeal to users about a certain topic, for example, "*I was totally fascinated again by Audrey's beautiful upright figure when I saw her on screen.*" In the **Standard** algorithm, the top 30% of sentences consisting of fifteen to twenty morphemes, sorted beforehand by TF-IDF scores, are extracted. Like the **Short** algorithm, the biggest TF-IDF is employed if a sentence has more than two nouns, and then the top 30% list is sorted by adjective frequency.

We assume sentences extracted by the **Diverse** algorithm express opinions or reasons with novel style, which can be unpredictable or sometimes serendipitous to users about a certain topic; for example, "*Roman Holiday is a romantic and sentimental story, which cute Audrey and gentle Gregory wove gorgeously.*" In this case, we assume the sentence should be long enough to express to author's opinion and its reasons to receive sympathy from interlocutors. With the unexpected expressions, those sentences are expected to attract a user's interest and make the conversation fun. In the **Diverse** algorithm, first, the top 30% of sentences consisting of fifteen to twenty morphemes, sorted beforehand by TF-IDF scores, are extracted. Next, the top 30% list is sorted in inverse order of adjective frequency.

The results are from the example "Castle in the Sky." The 2073 sentences are provided by extracting only positive opinions from "Castle in the Sky." 622 sentences are employed as the top 30% in terms of

Table 4.3: Example sentences from "*Castle in the Sky*" after sentence style conversation

子供の頃に見て，シータのあの青い石のペンダントが欲しかったよね.
(I wanted the blue stone pendant when I watched it as a child. Didn't you want it too?)
何回観ても面白いよね.
(No matter how many times I watch it, it is still fun, isn't it?)
宝探しという冒険を現代の子供たちは思い描く事が出来るんだ
(Kids today can imagine what treasure-hunting adventures are like.)

Table 4.4: Example sentences in the "*Castle in the Sky*" sorted by TF-IDF of nouns

Sentence	# of morph.	TF-IDF
パズーとシータがラピュタに到達したときの壮大な音楽が特に素晴らしいと思うんだ (The grandiose music was especially great when Pazu and Princess Sheata arrived at the**Castle in the Sky**.)	20	0.088
ラピュタはジブリ作品の中でも特に音楽が綺麗で素晴らしいと思うんだ. (The music of the **Castle in the Sky** is especially great and beautiful in Ghibli's movie.)	16	0.088
純粋で純真な心を持ったシータとパズー海賊の皆さんもとても可愛くチャーミングでピュアだよね. (Dola's family and **Princess Sheata** with pure mind are really cute, charming, and innocent.)	20	0.073
パズーのタフさに惚れるんだ. (We admire the toughness of **Pazu**.)	8	0.072
ジブリ作品で一番好きな映画だよ. (This is my most favorite movie from **Ghibli**.)	8	0.037
すべてにおいて宮崎駿の好きなものをちりばめた宝箱のような作品だよ. (This movie is like a jewel box containing the favorite stuff of **Mr. Miyazaki**.)	20	0.035
後押しするドーラ一家がまた素晴らしいよね. (**Dola's** family are great at boosting him.)	8	0.034

TF-IDF. Table 4.5, Table 4.6, and Table 4.7 show an example of each strategy. Bold adjectives are used for sorting the adjective TF.

4.4 System Architecture

We describe the architecture of our system for informative question answering based on the consideration in Section 4.2, and depicted in Fig. 4-4. The main processes in the framework are the natural language understanding (NLU) process, the dialogue management process, and the sentence generation process. The NLU process includes topic estimation and utterance (question) analysis. The sentence generation process is divided into factoid and non-factoid typed answer generation modules. The factoid typed answer generation module refers to structured knowledge databases organized using Semantic Web techniques. The non-factoid typed answer generation module generates the system's own opinions automatically extracted from a large indefinite number of reviews on the Web. The framework also has an utterance combination mechanism that combines factoid and non-factoid typed responses to realize the additional phrasing function.

4.4.1 Natural Language Understanding Process

In the NLU process, each spoken utterance or text input is interpreted with a current topic, a question type (5W1H interrogatives: e.g., "who," "what," "how," etc.) and a predicate (verbs and adjectives). We use a handcrafted dictionary for interpretation.

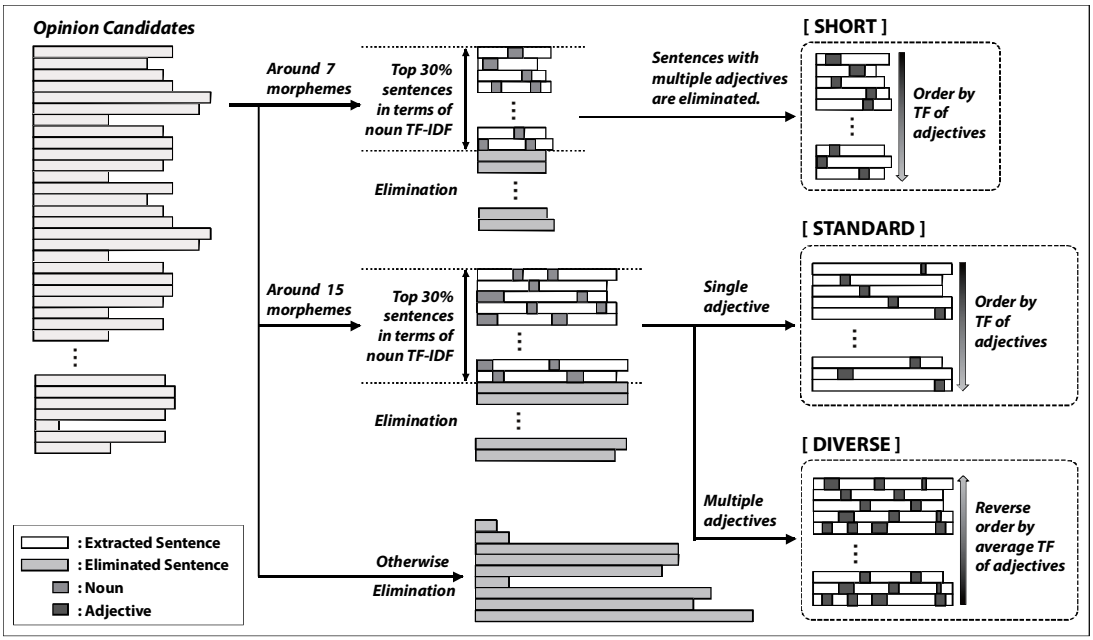


Figure 4-3: Ranking algorithms. After sentence candidates are sorted by TF-IDF scores, the top 30% of sentences consisting of approximately seven and fifteen morphemes are extracted, respectively. In the **Short** and the **Standard** algorithms, the lists are sorted by adjective frequency. In the **Diverse** algorithm, the list is sorted in the inverse order by adjective frequency.

In this paper, we define a sequence of topic words as a conversational context. Each system utterance is hooked to one of the topics. For example, the sentence "*Audrey is beautiful, isn't she?*" is assumed to belong to the topic "Audrey Hepburn." In our experimental system, we manually selected 100 topic words of the *movies* domain, which include popular titles, directors, and actors.

The topic estimation procedure uses the following three processes: Japanese language morphological analysis, important words filtering, and classification. After an automatic speech recognition (ASR) or text input is processed by Japanese language morphological analysis, only nouns are extracted. Then, the important nouns in each topic are extracted. In terms of degrees of importance, we use the TF-IDF score for each topic. We employed the top 50 important words for each topic. In the classification process, we used the linear-chain conditional random fields (CRF) technique. The following features are used in the prediction process:

$$topic_{t-2}, topic_{t-1}, topic_t, topic_{t-2}\&topic_{t-1},$$
$$topic_{t-1}\&topic_t, topic_{t-2}\&topic_{t-1}\&topic_t$$

where $topic_t$ denotes the topic word at time t. "&" symbol indicates a conjunction features. As an evaluation experiment, we evaluated the accuracy of 10-topic classification. We recorded three-minute conversations with two participants in which they were instructed to talk within 10 topics in the animation film domain. We conducted a total of 25 sessions. 20 of which were used for learning data, and five used for test data. The result for the accuracy rate (number of correct answers / total number estimated) was 74.8%

Table 4.5: Example sentences of the "*Castle in the Sky*" ranked by **Short** algorithm (adjectives are shown with boldface).

Sentence	Adj. TF
天空の城ラピュタが一番好きだよ. ("*Castle in the Sky*" is my most **favorite** movie.)	0.052
ドーラは一番好きなおばあさんだよ. (Dola is my most **favorite** woman.)	0.052
後押しするドーラ一家がまた素晴らしいよね. (Dola's family are **great** at boosting him.)	0.036
やっぱムスカのが一番すばらしいよね. (Anyway, Colonel Muska is **great**.)	0.036
宮崎映画で最高の作品だよ. (This is the **greatest** movie from Ghibli.)	0.029
特にドーラおばさんのキャラが最高だよね. (Captain Dola is the **greatest** character in this movie.)	0.029
ムスカ様の名言も面白いよね. (Wise Colonel Muska is also **interesting**.)	0.024

Table 4.6: Example sentences of the "*Castle in the Sky*" ranked by **Standard** algorithm (adjectives are shown with boldface).

Sentence	Adj. TF
ストーリーを通してあふれるパズーの命がけでシータを守る姿がいいよね. (It's **nice** that Pazu helps princess Sheeta at the risk of his life throughout the whole story.)	0.099
この映画をみてパズーのような男の子がいっぱい増えたらいいなあと思うんだ. (When I watch this movie, I wish there were a lot of boys like Pazu.)	0.099
これが一番宮崎駿作品の中で好きだよ. (This is my most **favorite** movie from Ghibli.)	0.052
宮崎駿作品で一番好きどころか，日本アニメ壱の名作だよ. (This movie is one of the best animation movies in Japan, as well as my **favorite** from Ghibli.)	0.052
ラピュタというものを通して人間の本質を描いた素晴らしい作品だよ. This movie is **great** at depicting the essence of human nature.	0.036
エンディングで流れる君をのせてやはり初期ジブリ作品はすばらしいよね. The song "Carrying You" at the end of this movie is **great** and the beginning of "Ghibli"'s series is awesome.	0.036
特にシータが飛空挺から落ちるシーンでのオープニングタイトルの入り方が最高だよね. The scene in which Princess Sheata is falling from the flying boat is the **greatest** at the start of this movie.	0.029

under a word error rate for ASR of 0%, and 65.2% under a word error rate for ASR of 20%.

4.4.2 Sentence Generation and Combination Process

The sentence generation process consists of two components: factoid-typed answer generation and opinion generation (non-factoid-typed answer generation). As an additional phrasing capability studied in Section 4.2, every factoid typed answer sentence always combines an additional opinion. On the basis of the result of the Question Analysis process, answers are classified into two types: factoid typed answers and non-factoid typed answers. Factoid answers are generated from a structured database. In our research, we used Semantic Web technologies. After analyzing a question, it is interpreted as a SPARQL query, a resource description framework (RDF) format query language to search RDF databases. We used DBpedia[8] as the RDF database (Auer et al., 2007).

[8]http://ja.dbpedia.org/

Table 4.7: Example sentences of the "*Castle in the Sky*" ranked by **Diverse** algorithm (adjectives are shown with boldface).

Sentence	Means of adj. TF
純粋で純真な心を持ったシータとパズー海賊の皆さんもとても可愛くチャーミングでピュアだよね. (Dola's family and Princess Sheata with **pure** mind are really **cute**, **charming**, and **innocent**.)	0.004
料理が上手く，上品で，優しく，愛らしく，時にはとても勇敢なシータだよ. (The brave Princess Sheata is **polished**, **tender**, **sweet**, and can cook **well**.)	0.004
パズーの勇気と強さシータのかわいらしさとやさしさどれをとっても最高だよね. (This movie is the best, with brave and **tough** Pazu, and **tender** and **cute** Princess Sheata.)	0.012
シータのしとやかで知的で可愛らしい，理想的な女の子像には，今も憧れるんだ. (I admire Princess Sheata with her **ladylike**, **cute**, and **intelligent** characteristics.)	0.001
決してスマートでカッコいい物に描いていない点が宮崎流だよ. (Mr. Miyazaki **never** created this movie **smart** and **stylish**, and that is good.)	0.004
ラピュタはすごい独自の世界観個性的なキャラクターそしてなによりシータとパズーが素敵だよね. ("Castle in the sky" has many **great unique** characters and Pazu and Princess Sheata are **sweet**.)	0.011
パズーの真っ直ぐさと，シータの純真さに自ずと涙が出てくるんだ. (I shed tears with **tame** Pazu and **pure** Princess Sheata)	0.002

4.4.3 Factoid-typed Sentence Generation

For factoid typed sentence generation, we employ Semantic Web techniques, which are widely used in question answering systems, such as in IBM's Watson (Ferrucci et al., 2010). The structured data in the Semantic Web is built on the W3C's resource description framework (RDF)[9], a mechanism for describing resources on the Web to store, exchange, and use as machine-readable information. DBpedia is a project that has the objective of extracting structured content from the information created as part of the Wikipedia project[10]. The DBpedia project uses RDF to represent the extracted information. DBpedia extracts factual information from Wikipedia pages, allowing users to find answers to questions where the information is spread across many different Wikipedia articles. RDF facilitates the making of statements about resources in the form of "Subject" - "Predicate" - "Object" expressions. The subject denotes the resource, which is usually a web resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object. For example, one way to represent the notion "The director of *Roman Holiday* is William Wyler" in RDF is as the triple: a subject denoting "*Roman Holiday*," a predicate denoting "director," and an object denoting "William Wyler."

In the question analysis process, the type of the question for asking about the fact or description is determined. The fact typed question is a question requesting an exact information (e.g. "Who is the director of *Roman Holiday*?"). The description typed question is a question requesting the abstract of the movie (e.g "What kind of movie is it?"). We use hand-crafted classification rules using predicates and specific nouns, which we defined. Sentences containing predicates explicitly asking a fact, such as directors, plots, and scenarios, are deemed questions about the fact. Sentences containing a specific predicate requesting a description or only an interrogative (e.g. "What") are deemed questions about the description.

Next, candidate sentences are searched for using the result of the question analysis process. When the question is determined as asking about the fact, a query is generated to search for an object using both subject and predicate as keys. SPARQL is used for the query, and the candidate results are searched for in the DBpedia database. For example, a question about the topic "Roman holiday" could be "Who is the actress of this movie?", for which the query in SPARQL would be:

[9]http://www.w3.org/RDF/
[10]http://ja.dbpedia.org/

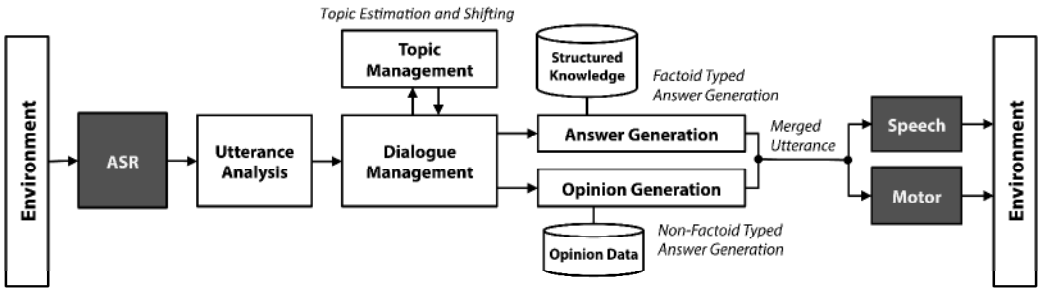


Figure 4-4: The main components in the architecture of the system are the Utterance Analysis, the Dialogue Management, and the Sentence Generation modules. The Utterance Analysis module receives sensory information from speech recognizers. The Dialogue Management module is described in Section 4.4. The Answer Generation module is capable of additional phrasing with the system's own opinions. The Opinion Generation process is described in Section 4.3.

```
PREFIX roman:
 <http://dbpedia.org/resource/Roman_Holiday>
PREFIX actors:
 <http://dbpedia.org/ontology/starring>
select distinct * where
 {roman: actors: ?actor .}
limit 10
```

When the questions are determined to be asking about the definition, the sentences asking about the abstract of the subject are set in the predicate and a query with the subject and the object as keys is provided for searching for the object.

```
http://ja.dbpedia.org/property/abstract
```

The abstract of "Roman holiday" is given below:

> *"Roman Holiday is a 1953 romantic comedy directed and produced by William Wyler. It stars Gregory Peck as a reporter and Audrey Hepburn as a royal princess out to see Rome on her own. Hepburn won an Academy Award for Best Actress for her performance; the screenplay and costume design also won."*

Since the sentences are too long to express the abstract of Roman Holiday in conversation, only the first sentence is regarded as an appropriate sentence and extracted. The various questions below possibly correspond with DBpedia:

- specific question: "What is a typical romance movie?"
  "Romeo and Juliet is one of those."

- complicated question: "How many people born after 1943 have won an Oscar?"
  "XXX people have."

Figure 4-7 shows acceptable predicates for the prototype system.

		Topic: "*Castle in the Sky*"
1:	User	Do you know "*Castle in the Sky*?"
2:	System	Yes, that movie is my favorite among "Ghibli" movies. **[Short]**
3:	User	OK, when was the movie released?
4:	System	Let me see, I think the movie was released in 1986.**[FactoidQA]**
5:	System	It's **nice** that Pazu helps princess Sheeta at the risk of his life throughout the whole story. **[Merged:Standard]**
6:	User	Yeah, I think so too.
7:	System	Pazu and his colleague have pure and innocent hearts, and they are really cute. **[Diverse]**
8:	User	That's right.
		Topic: "*Black Swan*"
9:	User	How about "*Black Swan*"?
10:	System	Natalie Portman is a good performer, I think. **[Short]**
11:	User	Yes, who is the director?
12:	System	Well, the director is Darren Aronofsky. **[FactoidQA]**
13:	System	The camera work in this movie is great, as well as Natalie's performance. **[Merged:Standard]**
14:	User	Yeah, right.
15:	System	It is an erotic thriller movie that also expresses the elegance of ballet. **[Diverse]**
16:	User	Oh, I didn't know that.

Figure 4-5: Examples of the system of in action. At #3, the user asks the system about a movie, then the system replies (#4). The system then adds an opinion related to the current topic (#5) during the same turn. A scenario with the same structure appears from #11 to #13.

4.5 Experiments

4.5.1 Experimental Design

We designed experiments focused on the quality and effectiveness of the utterance generation. For the purpose, one possibility would be to evaluate only by reading transcripts that had been automatically generated. However, there could be effectiveness that a robot speaks its own words as the first person. Therefore, in this paper, we designed both reading and video watching tasks. In the video watching tasks, an experimenter manipulated only the timings of a robot's responses along scenarios, where the robot's utterances were automatically generated.

We designed three experiments to evaluate (1) acceptability of sentences, (2) effectiveness of additional phrasing, and (3) effectiveness of sentence generation algorithms. In the first experiment, in order to evaluate only acceptability of each sentence, subjects were requested to read and rate each sentence extracted and ranked as we described in Section 4.3. In both the second and third experiments, each subject was requested to watch videos filmed about conversations between the robot and an interlocutor (person), and imagine as if he/she was the interlocutor talking with the robot. Fig.4-8 shows scenes from the videos. As shown in the figure, the camera was positioned in the back of a person and focused on the robot's face over the person's shoulder. This camera angle was intended that a subject could easily imagine he/she was an interlocutor of the robot. In order to maintain the same response timing, we did not use an ASR, but remotely manipulated the robot along scenarios as it replied to the person's question. The contents of the utterance of the robot were selected from the top of lists which were automatically generated as we described in Section 4.3.

Genre	Hollywood movie	Japanese movie	Directors and Actors
Movie	Indiana Jones	Bayside Shakedown	Harrison Ford
Foreign	Pirates of the Caribbean	Gantz	Steven Spielberg,
Japanese	Spiderman	Umizaru	Johnny Depp
Animation	Terminator	Trick	Tobey Maguire
Action	Roman Holiday	Juvenile	Arnold Schwarzenegger
Romance	Titanic	Always	James Cameron
Fantasy	Harry Potter	Koizora	Audrey Hepburn
Documentary	The Lord of the Rings	Water Boys	Leonardo DiCaprio
Sci-Fi	Alice in Wonderland	Kaiji	Daniel Radcliffe
Comedy	Super Size Me	Departures	Tim Burton
Suspense	Men in Black	Ring	Will Smith
Horror	Back to the Future	One Missed Call	Chris Columbus,
Spectacular	Home Alone	The Wow-Choten Hotel	Bruce Willis
Sports	SAW	Summer Wars	Bruce Lee
Panic	The Sixth Sense	Pokemon	Michael Bay
Mistery	Cube	My Neighbor Totoro	Yuji Oda
Hollyhood	Enter the Dragon	Castle in the Sky	Kazunari Ninomiya
Ghibli	Paranormal Activity	Whisper of the Heart	Kenichi Matsuyama
Disney	Red Cliff	Princess Mononoke	Hiroshi Abe
Actor	Shaolin Soccer	Spirited Away	Yukihiko Tsutsumi
Director	Jaws	Cinderella	Tatsuya Fujiwara
	Armageddon	Beauty and the Beast	Motoki Masahiro
	Transformers	Toy Story	Nanako Matsushima
			Ko Shibasaki
			Koki Mitani
			Mamoru Hosoda
			Hayao Miyazaki
			Gregory Peck
			Orlando Bloom
			Keira Knightley

Figure 4-6: Sample topics used in the current version.

4.5.2 Experimental Platform

We used the multimodal conversation robot "SCHEMA ([∫e:ma])" as our experimental platform (Matsuyama et al., 2009), which has fundamental abilities to follow conversation protocols (Kobayashi and Fujie, 2013). SCHEMA is approximately 1.2 m in height, which makes it level with the eyes of an adult male sitting in a chair. It has 10 degrees of freedom for right-left eyebrows, eyelids, right-left eyes (roll and pitch) and neck (pitch and yaw). It can express anxiety and surprise using its eyelids and control eye gaze using eyes, neck, and autonomous turret. It also has six degrees of freedom for each arm, which can express gestures. One degree of freedom is assigned to the mouth to indicate explicitly whether the robot is speaking or not. A computer is inside the belly to control the robot's actions, and an external computer sends commands to execute various behaviors via a WiFi network. All modules, including the ASRs and a speech synthesizer, are connected to each other through a middleware called the Message-Oriented NEtworked-robot Architecture (MONEA), which we produced (Nakano et al., 2006).

4.5.3 Experiment 1: Acceptability of Sentences

We conducted this experiment to evaluate acceptability of each sentence. A total of 5 male subjects participated in the experiment. All subjects were graduate school students with an average of 23 years old, who are native Japanese speakers recruited from Waseda University campus. The subjects were first given a brief explanation of the purpose of the experiment, and then they were requested to read and rate each sentence

Predicates	Resource URLs in RDF format
Starring	http://ja.dbpedia.org/property/出演者
Screenplay	http://ja.dbpedia.org/property/脚本
Director	http://ja.dbpedia.org/property/監督
Editing	http://ja.dbpedia.org/property/編集
Gross	http://ja.dbpedia.org/property/興行収入
Costume	http://ja.dbpedia.org/property/衣装
Priduce	http://ja.dbpedia.org/property/製作
Budget	http://ja.dbpedia.org/property/製作費
Language	http://ja.dbpedia.org/property/言語
Music	http://ja.dbpedia.org/property/音楽
Academy Awards	http://ja.dbpedia.org/property/アカデミー賞
Emmy Award	http://ja.dbpedia.org/property/エミー賞
Grammy Awards	http://ja.dbpedia.org/property/グラミー賞
Golden Globe Awards	http://ja.dbpedia.org/property/ゴールデングローブ賞
Tony Award	http://ja.dbpedia.org/property/トニー賞
British Academy Film Awards	http://ja.dbpedia.org/property/英国アカデミー賞
Genre	http://ja.dbpedia.org/property/ジャンル
Filmography	http://ja.dbpedia.org/property/主な作品
Place of birth	http://ja.dbpedia.org/property/出生地
Other names	http://ja.dbpedia.org/property/別名
Family	http://ja.dbpedia.org/property/家族
Real nane	http://ja.dbpedia.org/property/本名
Birth date	http://ja.dbpedia.org/property/生年月日
Death date	http://ja.dbpedia.org/property/没年月日
Years active	http://ja.dbpedia.org/property/活動時期
Occupation	http://ja.dbpedia.org/property/職業
Height	http://ja.dbpedia.org/property/身長
Spouse	http://ja.dbpedia.org/property/配偶者

Figure 4-7: Sample predicates used in the current version.

in lists extracted and ranked just like Table 4.5 to 4.7. We selected the latest 5 popular movie titles as topics ("*The Tale of Princess Kaguya*", "*Gravity*", "*World War Z*", "*Eien no Zero*" and "*The Wind Rises*"), whose number of reviews are large enough to extract opinion sentences for this experiment.

We created 12 conditions in terms of ranking algorithms (Short, Standard, Diverse) and topic coherence (the TF-IDF scores). As the parameter of the topic coherence, we controlled the threshold of the average of noun TF-IDF scores, which we described in Section 4.3.4 (we employed top 30% in the section). In this experiment, we employed four thresholds (Top 10%, 30%, 50% and 70% sentences) for each algorithm. The overlap rate among four topic coherence thresholds, which was calculated as content rate of sentences of top 10% in top 30%, 50% and 70% sets, averaged 39.1%, 22.2% and 16.2%, respectively. Each condition contains 30 sentences. Totally 1,800 sentences (12 conditions $\times$ 5 topics $\times$ 30 sentences) were evaluated by each subject.

We defined a metric of acceptability as a combination of grammatical appropriateness and subjects' impression of topic coherence of each sentence, which means if a sentence is evaluated as grammatically appropriate and coherent to a certain topic at a same time, the sentence would be acceptable. Each subject was requested 3-scale Likert questionnaires to evaluate each sentence in terms of the grammatical appropriateness ("3" is "appropriate ", "2" is "not sure" or "conditionally appropriate", and "1" is "not appropriate") and the topic coherence. As for the topic coherence, subjects were instructed as follows: "If you can clearly judge a sentence is describing a certain aspect of a current topic, then rate it 3. If you are not sure whether a sentence is describing a certain aspect of a current topic or not, then rate it 2. If you can clearly judge a sentence is not describing a certain aspect of a current topic, then rate it 1". If both the grammatical

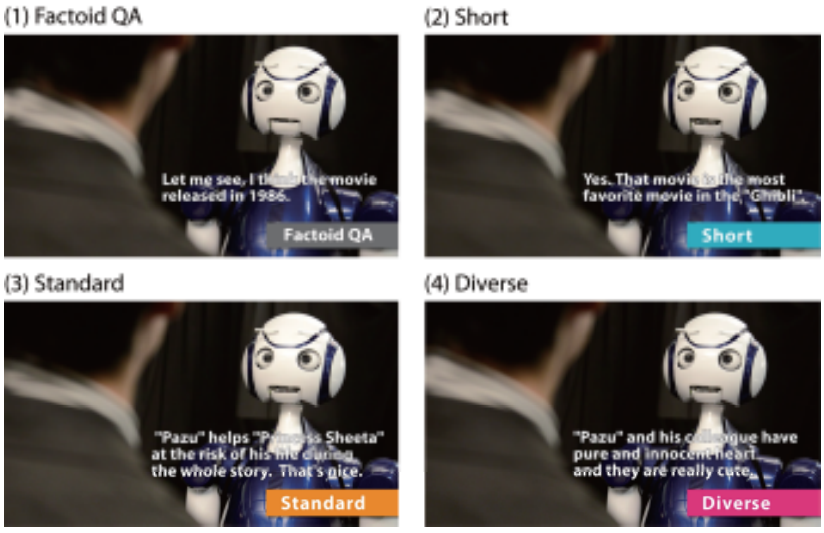


Figure 4-8: Sample scenes of different sentence generation algorithms: (1) Factoid typed answer, (2) Short typed opinion, (3) Standard typed opinion, and (4) Diverse typed opinion. Subtitles were not included in the videos previewed in Experiment 2 and 3.

appropriateness and the topic coherence are rated as "3", acceptability is rated as "acceptable", otherwise, rated as "unacceptable".

4.5.4 Results and Discussion of Experiment 1

Fig. 4-9, 4-10 and 4-11 show the result of the grammatical appropriateness, the topic coherence and acceptability, respectively. We conducted the analysis of variance (ANOVA) in terms of algorithms. There are significant differences among algorithms in all metrics ($p<0.01$). We also conducted the ANOVA in terms of the TF-IDF scores. On the grammatical appropriateness, there were no significant differences. The result shows the TF-IDF scores don't significantly affect the grammatical appropriateness. On the topic coherence, significant differences were found in Standard ($p<0.03$) and Diverse algorithms ($p<0.01$), where there are trends to gradually decrease. On the topic acceptability, significant differences were found only in Diverse algorithm ($p<0.03$).

As for the topic coherence, although it would be reasonable that we employ top 10% of sentences in terms of TF-IDF according to the result, some topics have few sentence candidates in the top 10%, which depend on review resources. Therefore, we employ top 30% in order to extract enough number of sentences in the following experiments. As for acceptability, an interesting finding was that there is a small trend increasing between 50% and 70% in Diverse algorithm ($p<0.3$). We can not conclude this is a significant trend, however, we might be able to hypothesize that subjects are not so concerned about a certain level of diversity of topic coherence and grammatical mistakes in Diverse sentences, which should be verified in further experiments. Overall, the automatically extracted and ranked opinions we proposed got 78.8%, 75.7% and 73.6% of acceptability in Short, Standard and Diverse algorithms, respectively, using the top 30% of sentences. These results also show that there are possible ways to dynamically control these parameters (TF-IDF, length of sentences, and adjective TF) to extract and rank sentences to get as many

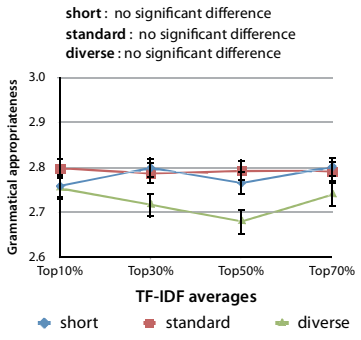


Figure 4-9: Grammatical appropriateness.

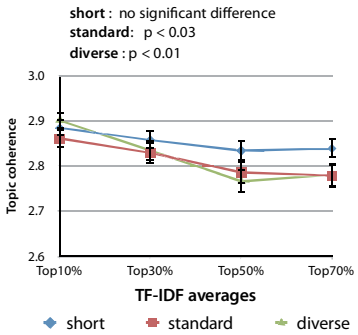


Figure 4-10: Topic coherence.

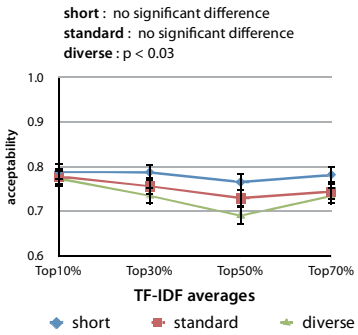


Figure 4-11: Acceptability.

good sentences as possible.

4.5.5 Experiment 2: Additional Phrasing

The second experiment was designed to evaluate additional phrasing functions. Two conditions were videotaped, and all videos contained the same two topics ("*Castle in the Sky*" and "*Black Swan*"). As shown below, scenarios of the condition 1 and 2 were lexically identical, except that the condition 2 had additional phrasings.

- **Condition 1 (Simple Answering)**: A simple question answering system that replies to user's question simply; for example, when a user asks "Do you like *Castle in the Sky*?", the system replies "Yes, I do."
- **Condition 2 (Additional Phrasing)**: The system replies with combined sentences: a simple answer and a related opinion. For example, when a user asks "Do you like *Castle in the Sky*?", the system replies "Yes, I do. It's nice that Pazu helps princess Sheeta at the risk of his life throughout the whole story."

The sentences were generated only using the **Standard** algorithm. An excerpt from the transcript of the experiment (condition 2) is shown in Fig. 4-12. A total of 32 subjects (21 males and 11 females) participated in the experiment. All subjects were native Japanese speakers recruited from Waseda University campus. The ages of the subjects ranged between 20 and 30 years, with an average age of 20.5 years. The subjects were first given a brief explanation of the purpose and the procedure of the conversation by a document and an oral presentation, which includes the following: "The purpose of this experiment is evaluation of the ways of expressions of robot's utterances themselves (e.g. length of the sentence, and diversity of vocabulary). Please ignore other factors, such as its intonation, timing and variety of topics. ... Imagine that you were the interlocutor of the robot, and you were interested in these topics from the beginning." And the subjects could ask any questions about the experimental setting of the experimenter, so that every subject clearly understand the situation. After they watched the videos, they were asked to give evaluations about Enjoyment ("Which condition did you feel was more enjoyable?"), Politeness ("Which condition did you feel was more polite?"), and Personality ("Which condition did you feel had a better personality?"). Subject could also select "No differences" for each question and answer further comments in free-forms.

User	Do you know "*Castle in the Sky*?"
System	Yes, I do. **[Minimum Response]**
	It's nice that Pazu helps princess Sheeta at the risk of his life throughout the whole story. **[Merged Opinion]**
User	Do you remember the director of the movie?
System	It's Hayao Miyazaki. **[Minimum Response]**
	Every Miyazaki movie is wonderful, isn't it? **[Merged Opinion]**

Figure 4-12: Excerpt from the transcript of Experiment 2. (Condition 2). Scenarios of the condition 1 and 2 are lexically identical, except that the condition 2 had additional phrasings.

4.5.6 Results and Discussion of Experiment 2

Figure 4-14 shows the result of the experiment. 72% of the subjects answered that they felt enjoyment with the proposed system (3% answered there is no difference), 63% answered that they felt politeness with the proposed system (3% answered there is no difference), and 59% answered that they liked the robot's personality with the proposed system.

Most of subjects felt enjoyment about the additional mechanism. However, in the free-form questionnaires, some subjects reported that they persistently felt the robot's personality. The result implies that the additional function should be switched over based on conversational contexts.

4.5.7 Experiment 3: Comparison of Sentence Generation Algorithms

We conducted this experiment to evaluate effectiveness of sentence generation algorithms using recorded video watching. We used the following three evaluation metrics: (1) users' impressions of enjoyment, (2) users' motivations for participation, and (3) users' impressions of the robot's personality.In terms of users' impressions of the enjoyment metrics, users were asked if they felt enjoyment from a conversational sequence of a video they watched. Users' motivations for participation examined how users felt to participate in the conversation. Users' impressions of the robot's personality examined how users had positive impressions of the robot's personality. The videos filmed about conversations between the robot and a person were shown. The video was watched by the subjects, after which questionnaires were given to them. A total of 38 subjects (21 males and 17 females) participated in the experiment. All subjects were native Japanese speakers recruited from Waseda University campus. The ages of the subjects ranged between 20 and 30 years, with an average of 22.0 years. The subjects were first given a brief explanation of the purpose of the experiment, which is same as the experiment 2, and then they watched the following four types of conversations:

- **Condition 1 (Short)**: The robot outputs six **Short** sentences.
- **Condition 2 (Standard)**: The robot outputs six **Standard** sentences.
- **Condition 3 (Diverse)**: The robot outputs six **Diverse** sentences.
- **Condition 4 (Mixed)**: The robot outputs two sentences each for **Short**, **Standard**, and **Diverse**, respectively.

The videos comprised four different situations. Fig.4-8 shows one scene from the videos. Each one-minute-long video includes three topics, where a robot outputs two sentences for each topic. Four genres were used, with each genre containing the following three topics:

	Topic: "*Ghibli's*"
User	Let's talk about "*Ghibli's*"
System	Yes.
User	Do you know "*Spirited Away*" ?
System	Yes, it is a good movie because we can see Chihiro's growth. **[Short]**
User	Yeah, I think so too.
System	I love it the best among Ghibli's. **[Short]**
User	I see. So, do you know "*From Up on Poppy Hill*" ?
System	Yes, I think the movie speaks to Goro's (the director) good-naturedness. **[Standard]**
User	Right.
System	It is a nice movie set in Yokohama in the 1960's. **[Standard]**
User	Yeah. So, do you know "*Castle in the Sky*"?
System	Yes, Pazu and his colleague have pure and innocent hearts, and they are really cute. **[Diverse]**
User	I see.
System	I still love her ideal image as a modest and intellectual girl. **[Diverse]**

Figure 4-13: A sample transcript from Experiment 3 (Topic: "*Ghibli*," Condition: "Mix").

1. Ghibli's ("*Spirited Away*" → "*From Up On Poppy Hill*" → "*Castle in the Sky*")
2. Animation movies ("*SUMMER WARS*" → "*ONE PIECE*" → "*EVANGELION*")
3. Actors ("*Yuji Oda*" → "*Tatsuya Fujiwara*" → "*Ken-ichi Matsuyama*")
4. Hollywood movies ("*Back to the Future*" → "*Inception*" → "*Black Swan*")

We set up the following eight types of algorithm orders to cancel order effects:

1. Ghibli's (**Short**)→Animation (**Standard**)→Actors (**Diverse**)→Hollywood (**Mixed**)
2. Ghibli's (**Standard**)→Animation (**Diverse**)→Actors (**Mixed**)→Hollywood (**Short**)
3. Ghibli's (**Diverse**)→Animation (**Mixed**)→Actors (**Short**)→Hollywood (**Standard**)
4. Ghibli's (**Mixed**)→Animation (**Short**)→Actors (**Standard**)→Hollywood (**Diverse**)
5. Ghibli's (**Mixed**)→Animation (**Diverse**)→Actors (**Standard**)→Hollywood (**Short**)
6. Ghibli's (**Diverse**)→Animation (**Standard**)→Actors (**Short**)→Hollywood (**Mixed**)
7. Ghibli's (**Standard**)→Animation (**Short**)→Actors (**Mixed**)→Hollywood (**Diverse**)
8. Ghibli's (**Short**)→Animation (**Mixed**)→Actors (**Diverse**)→Hollywood (**Standard**)

A sample transcript from Experiment 3 (Topic: "*Ghibli*," Condition: "Mix") is shown in Fig.4-13. The following questionnaire was also given to the subjects to choose the videos from our proposal. "Which is the most enjoyable answer in those videos?" (enjoyment), "Which is the most attractive answer willing you to join in those videos?" (motivation), and "Which personality of the robot is your favorite in those videos?" (personality)

4.5.8 Results and Discussion of Experiment 3

The results are shown in Fig. 4-15. Most of the people chose anything besides the Short. For the question of enjoyment, the same number of people chose Standard and Diverse. For the question of motivation and personality, the number of people who chose Diverse was twice as many as those who chose Standard.

We assumed that sentences that were not short but had certain information resulted in an enjoyable conversation. The opinions from the subjects, who gave evaluations containing everything but Short, indicate that the length of the sentences is appropriate. We also confirmed from the results of the questionnaires that all lengths apart from Short are suitable. While the number of people divided between Standard and Diverse in the question for enjoyable was the same, there were some opinions suggesting that the expression of Diverse was unexpected and attractive, such as "*The robot explained his opinion of the movies in detail and was specific*," "*I had fun hearing his opinion, I felt that his opinion was unique*," "*He expressed his feeling like he actually watched the movie*" and "*The expression of his feeling was the most expressive for those movies*." We presume that this rich expression attracted the mind of the subjects, making them expect more. Let us now examine the reason for the increased favorable rating caused by the Diverse sentences. The Diverse sentences have very rich expressions with details and specific descriptions, which we presume to give the impression that the information was not simply read from the Web. Consequently, the Diverse utterances attracted users' interests and made them willing to talk, making the robot more favorable.

4.6 Conclusions and Future Work

4.6.1 Summary and Contributions

We presented automatic sentence generation mechanisms for enjoyable conversational systems, including expressive opinion generation and additional phrasing mechanisms. Its opinion sentences are generated from a large number of reviews present on the web. After it conducts an opinion extraction and sentence style conversion process, opinion candidates are ranked in terms of contextual relevance, length of sentences, and frequency of adjectives. We conducted three controlled lab experiments to evaluate acceptability of sentences, and effectiveness of the opinion generation and additional phrasing mechanisms. In the first experiment, we evaluated acceptability of generated opinion sentences only by reading texts. The result shows acceptability averaged around 75%. In the second and third experiments, each subject was requested to watch videos and imagine as if he/she was an interlocutor of a robot in order to evaluate the robot's spoken opinions. Videos were filmed about conversations between a robot and a person, where the camera was positioned in the back of a person and focused on the robot's face over the person's shoulder. While the subjects were not real users, we assume that results could reasonably imply generated sentences themselves have potentials to promote real users' enjoyment and interests.

The main contribution of this research to the question answering systems research domain is considerations of informative responding, specifically, expressive opinions. Conventional question answering systems are primarily focused on functional interactions to achieve specific tasks; therefore, they are based on Grice's cooperative conversation principles. However, they are not enough to attract users to engage with the system. In this paper, we assumed informative productions in our daily enjoyable conversations appear as an interlocutor's original way of expressions and viewpoints, which can be represented as frequency of adjectives. Beyond simple exchange of questions and answers, just like most current academic and industrial question answering systems, the expressive opinions and additional phrasing mechanisms could possibly trigger users' motivations to continue to interact with systems over a period of time. In the

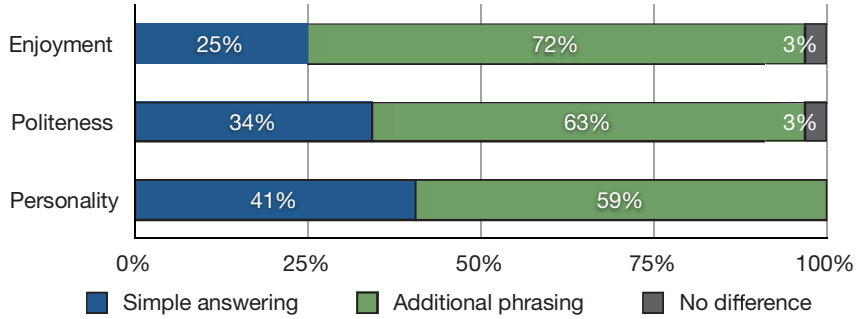


Figure 4-14: Result of Experiment 2: Impression of additional phrasing.

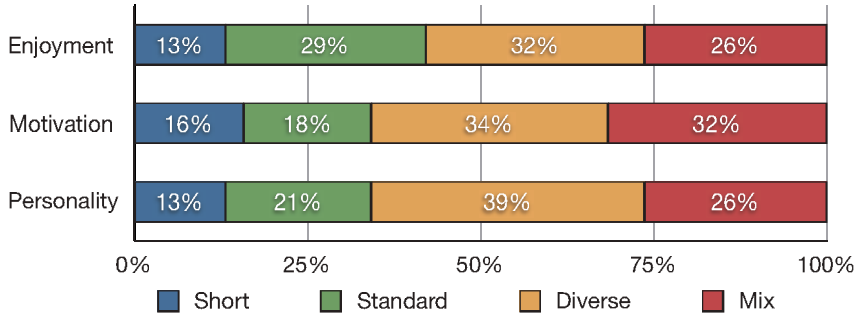


Figure 4-15: Result of Experiment 3: Comparative results of ranking algorithms.

following subsections, we discuss further extensions of this research.

4.6.2 Contextual Tracking

When a user follows up with the system about an opinion, he/she would be motivated to talk deeper about the system's opinion. Our current experimental system can only track a sequence of topics and select a sentence from the top of the list. The future works should include considering extending the model of conversational context and its tracking mechanisms. One possible way is to employ sophisticated topic models. In that domain, many statistical models, such as the probabilistic Latent Semantic Analysis (pLSA) (Hofmann, 1999), the Latent Dirichlet allocation (LDA) (Blei et al., 2003), have been proposed. Although these methods are suitable for discovering the abstract topics in written documents, there are few cases to apply to interactive dialogue systems. Another possible way is to estimate discourse structures. Grosz et al. and Walker et al. proposed and discussed the Centering theory (Grosz et al., 1995; Walker et al., 1998) to model the local coherence of discourse. Based on feasible contextual tracking mechanisms, we will conduct experiments evaluating enjoyment of conversations between a robot and real users, where they ask questions as they want.

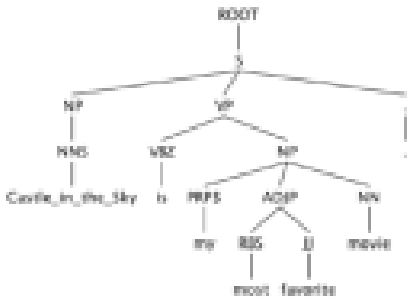


Figure 4-16: Syntax tree: "Castle in the Sky is my most favorite movie." (**Short**)

4.6.3 Syntactic Structure Control

Mairesse et al. proposed PERSONAGE, a highly parametrizable sentence generator that can manipulate parameters in terms of extraversion and introversion of personalities (Mairesse and Walker, 2007). Based on the psychological finding that introverts produce more complex constructions (Furnham, 1990), PERSONAGE can control syntactic structure complexity (depth of syntactic structures) in its syntactic templates selection stage. In this chapter, while we only controlled the length of sentences with respect to syntactic structure, it could eventually reflect syntactic complexities. Figure 4-16, 4-17 and 4-18 show examples of parsed syntax trees. However, this method does not guarantee reliable result of syntactic structure generation. We will consider more sophisticated syntactic template selection methods.

4.6.4 Recommendation with Expressive Opinions

Our proposed informative question answering system framework has huge potential for QA-typed recommender systems. Misu et al. presented a system-initiative information recommendation method as an application of a question answering system for user satisfaction (Misu and Kawahara, 2007). They proposed system-initiated spontaneous questions as a recommendation method. Beyond that, a mechanism of our system's additional expressive opinion has potential to offer serendipitous ideas to users, which might realize more enjoyable conversations. Ziegler et al. achieved high user satisfaction ratings by making topic-diversified recommendations and reducing the similarity of recommendation lists (Ziegler et al., 2005). Murakami et al. assumed that user satisfaction depends on whether a recommender system suggests unexpected items that are relevant to the users' preferences (Murakami et al., 2008). In order to apply our framework for more serendipitous recommender systems, we will consider learning mechanisms of opinion ranking algorithms and sentence combination based on conversational contexts and user models to maximize users' satisfaction.

4.6.5 Application to Other Domains

The opinion generation routine we implemented can generally be applied to other domains. We have applied it experimentally to domains such as *travel* and *foods*, which are typical "safe" small talk topics (Holmes, 2000). In the *travel* domain, we collected review documents from the 4travel travel review site[11]. Currently,

[11] http://4travel.jp/

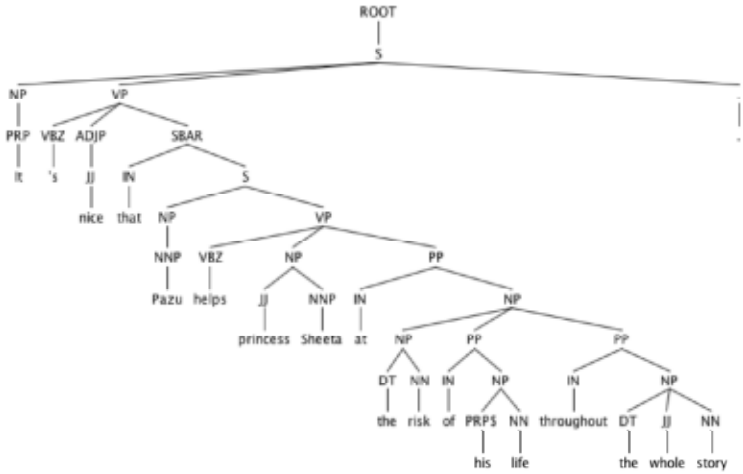


Figure 4-17: Syntax tree: "It's nice that Pazu helps princess Sheeta at the risk of his life throughout the whole story." (**Standard**)

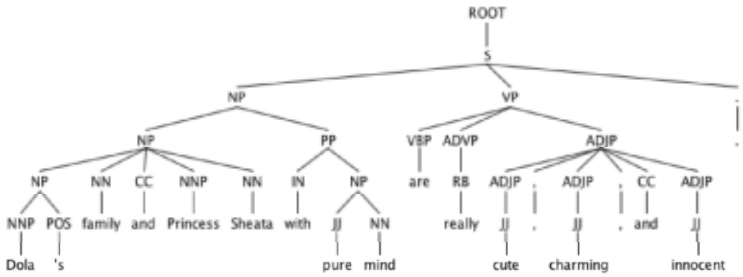


Figure 4-18: Syntax tree: "Dola's family and Princess Sheata with pure mind are really cute, charming, and innocent." (**Diverse**)

a large number of opinions about more than 30,000 travel spots (e.g., the Louvre Museum, the Eiffel Tower, and the Colosseum) have been generated. In the *foods* domain, we collected review documents from the Tabelog food review site[12]. Opinions about all restaurants in the Takadanobaba area of Tokyo (more than 1400 restaurants) have been generated and used for restaurant recommendation tasks. We decided to employ these review sites because of both the substantial volume and the quality of the reviews they have. These opinions can generally be used in question answering systems.

[12]http://tabelog.com/

"Anatomists draw these fundamental forms as "schema" in their minds. [···] Human anatomy, therefore, draws clearly such schemas of human bodies."

Shigeo Miki

5 SCHEMA: Robotic Platform

As a robotic platform suitable for multiparty conversation facilitation, we present the design of a robot called SCHEMA. In order to participate in multiparty conversational situations and be recognized as a ratified participant, a robot needs to have capabilities to exchange conversational protocols, which include organizing participation structure and transmitting messages. Such protocols essentially need robots' embodied functions such as facial expressions, head gestures, and directional control of torso. On the basis of our studies, SCHEMA is designed to be of approximately 120 cm tall and has 22 degrees of freedom, enabling the robot to perform essential conversational tasks. The robot is also designed with an user-friendly styling to appeal to users of all ages, from children to elderly people.

5.1 Introduction

In recent years, there has been a growing worldwide attention to the development of humanoid robots designed for social interactions, and such robots have been investigated in many fields, including human-robot interaction, developmental robotics, and embodied conversational agents. In such fields, robotic platforms have been regarded both as tools to model human cognitive functions from scientific perspectives, and as optimized human interfaces from engineering perspectives. While robotic hardware has been designed from both types of perspectives, recent robotic platforms have common architectural structures. The architectures can be defined in terms of protocols. Figure 1 shows a diagram of one of the common structures. The highest abstraction on the top is cognitive architecture such as a conversational system framework. Modules of a conversational system framework could run on multiple computer node located remotely. In this level, a higher level of protocols are defined to connect cognitive modules. Connections among the modules of a cognitive architecture are supported by networking middleware managing modularity by abstracting algorithmic modularity and hardware interfacing. The protocols of networking middleware address interprocess communication requirements. The abstraction is usually implemented as port connections. Some types of middleware follow the observer pattern by decoupling producers and consumers. Examples include

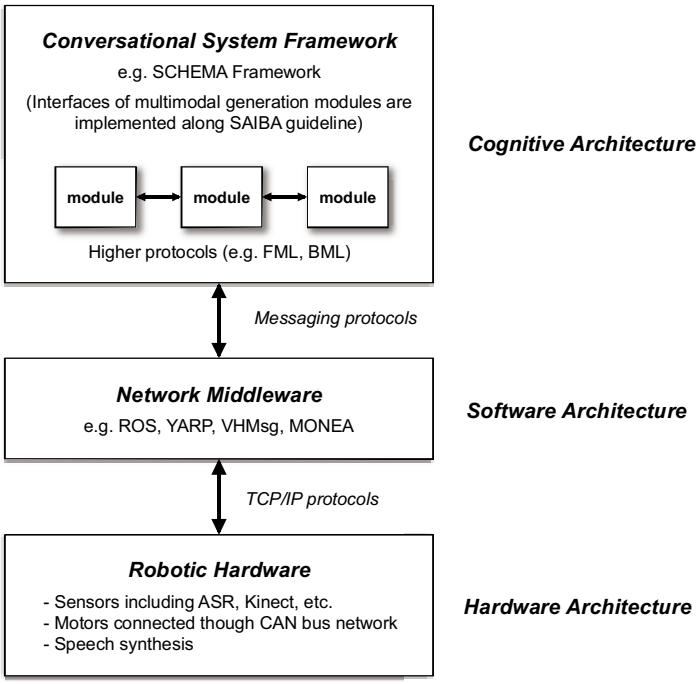


Figure 5-1: General layered architecture model of robotic platform. Modules of a conversational system framework (cognitive architecture level) could run on multiple computers that are located remotely and are supported by a networking middleware (software architecture level). Motor controllers and sensors are located on the robotic hardware. Communication among motor controllers and sensors is effected by a suitable connection protocol (e.g., CAN bus).

"yet another robot platform" (YARP) and the virtual human messaging (VHMsg) library. Such middleware can deliver messages of any size across a network using a number of protocols and shared memory. Some other middleware use peer-to-peer models such as the robot operating system (ROS) and message-oriented networked-robot architecture (MONEA). The lower abstraction level concerns hardware devices defining the interfaces for devices via their native APIs, which easily encapsulates hardware dependencies.

We present a robotic platform for conversational robots participating in multiparty conversations. In conversations among humans, we use social cues expressed by embodiments, such as facial expressions, head gestures, and body orientation. For example, head gestures such as nodding/shaking head can express positive/negative attitudes. Eye gaze explicitly expresses the participant's interest. ROBITA had capabilities to recognize the facial direction of the speaker (Matsusaka et al., 2003). When ROBITA detects the end of the speaker's utterance and that the speaker is facing the robot, it assumes that the turn is being handed over to it. It thus takes the turn and begins to speak. If it detects that the speaker is facing another hearer, it assumes that the turn is being handed over to that hearer. It regards that hearer as the expected speaker and gazes at that person (even if he or she does not begin to speak). This function achieves not only smooth turn-taking but also a feeling of unity. Besides, the robot's user-friendly exterior is necessary in the daily life situations where humans live with robots. Considering these perspectives, we developed SCHEMA, a robotic platform for physically situated conversational agents.

The platform was named SCHEMA ([$\int e$:*ma*]) because of the following quotation by the Japanese

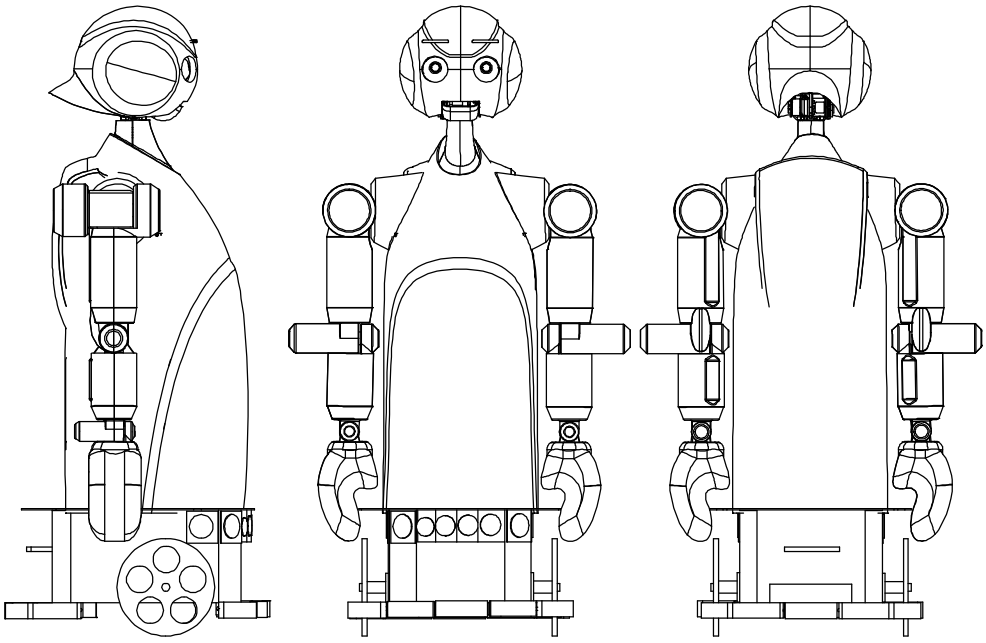


Figure 5-2: Exterior of SCHEMA

anatomist Shigeo Miki in his book “Introduction to Life Morphology”: ”Anatomists draw these fundamental forms as “schema” in their minds. Human anatomy, therefore, draws clearly such schemas of human bodies.” Generally, the word “schema” comes from the Greek word “$\sigma\chi\eta\mu\alpha$,” which means shape, or plan. As our platform is developed to investigate the general framework and process many aspects of conversations, we named it SCHEMA. SCHEMA is a successor of the previously developed “Waseda Robots” named ROBISUKE and KOBIAN. ROBISUKE is a one-meter-tall conversational robot developed by the Perceptual Computing Group in 2002, and numerous conversational systems have been developed based on the robot (Fujie et al., 2008). KOBIAN (Endo et al., 2008) is a full body bipedal walking humanoid with a face that enables it to express emotions. The robot was developed by combining and redesigning WABIAN, a bipedal robot (Ogura et al., 2006) and WE-4, an expressive-face robot (Miwa et al., 2002). Considering these concepts and mechanical designs, we redesign a humanoid robot for conversational purposes.

5.2 Exterior Design

Lee et al. (Lee et al., 2009) studied the body size of Snackbot, a robot designed to deliver snacks in a university building. The authors conducted an experiment using a prototype robot. The robot had three height conditions: small (112 cm), middle (128 cm), and high (142 cm) robots. They used a five-point Likert scale (1: much too small and 5: much too tall), to understand how friendly and intelligent people felt the robot was, and how they responded to the height of the robot. As the result of the Likert scale and free-form questionnaire, participants liked the tallest robot the most because they could make eye contact

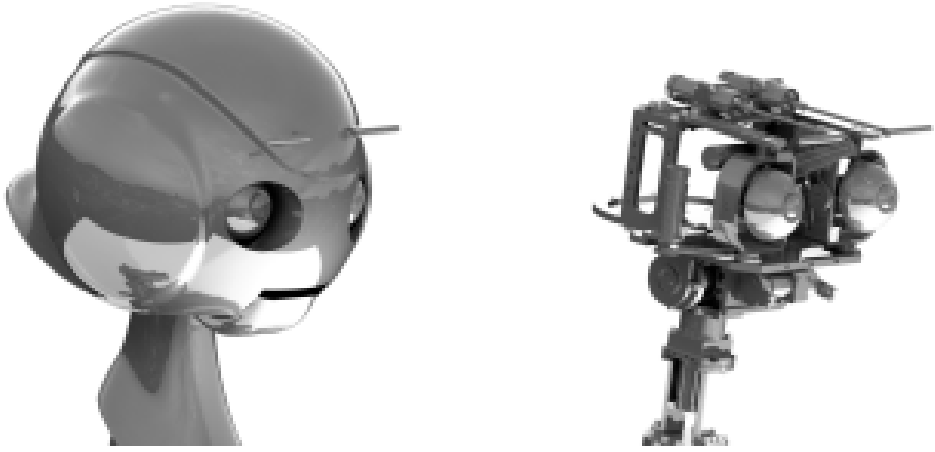


Figure 5-3: Head design of SCHEMA (mechanical - covered)

with it, and disliked to bend to interact with the smaller robots. In our study, taking into consideration both standing and sitting situations, we designed the height of the robot as 120 cm, slightly smaller than eye level. Ideally, it is the best way to implement all embodied capabilities in just proportion of humans to express social cues. Nevertheless, all degrees of freedom of the human body are not always needed to express social cues. We designed the degrees of freedom according to the priority of essential social cue capabilities. As for the styling (cover) design, we considered the following three functions:

1. Abstracted user-friendly forms.
2. Detachable joints of covers in case of emergency breakage because of high tension of wires.
3. Covers easily detachable by hand for maintenance.

To ensure user-friendly design, all parts are of the rounded streamline shape. In order to realize the free-form curve, we used the fiber reinforced plastics (FRP) material. As shoulders need bigger output motors, they might threaten users. As a solution, we crane its neck and treated surface treatment with free-form curves from head to shoulders to make the robot appealing to users. Considering safety and maintenance aspects, all joints of covers are attached to mechanical parts by strong magnets. Hence, screw holes are not required to attach the parts, resulting in beautiful styling.

5.3 Mechanical Design

For a socially situated conversational robot, head gesture and facial expression features are necessary. For example, head gestures such as nodding or shaking head are needed to express positive/negative attitudes. Eye gaze explicitly expresses the participant's interest in the conversation. Lip that move when the robot speaks and eyebrows that express emotions such as confusion and surprise are also needed. On an empirical basis of character animation, eye blinks are used to convey that the agent is "breathing." In order to realize these functions, the degrees of freedom of SCHEMA's head are designed as listed in Table 5.1. Arm movements are used for pointing at objects of interest or to communicate symbolic or linguistic information.

Table 5.1: Degrees of freedom for SCHEMA's head

Region	Degrees of freedom
Eyebrow	2 (roll axes)
Eyelid	1 (pitch axis)
Eyeball	4 (yaw and pitch axes)
Lip	1 (pitch axis)
Neck	2 (yaw and pitch axes)

Table 5.2: Degrees of freedom for each arm of SCHEMA

Region	Degrees of freedom
Shoulder	2 (pitch and roll axes)
Elbow	2 (yaw and pitch axes)
Wrist	2 (roll and yaw axes)

In order to realize these functions, the degrees of freedom of SCHEMA's arms are designed as listed in Table 5.2. Body orientation of a robot also generates social cues that elicit recognition by the participants in a group. Generally, participants place themselves where they can see each other and position their bodies so that they are oriented toward the group centroid. An overhearer willing to participate in the conversation typically gazes at the current speaker and expresses his/her intention to participate by certain actions such as raising a hand or saying a typical phrase, such as "excuse me..." If the overhearer is recognized by the current speaker, the other participants might change their positions so that they are oriented toward the centroid of the enlarged group. We design this function as the rotation of the mobile turret. In case the robot cannot follow an object enough only by eye gaze, the turret rotation can help cover the object. Motor drivers, laptop PCs, a speaker, a speaker amplifier, and batteries are located inside the robot's body. Table 3 lists the mechanical parts. Figure 5-4, 5-5, 5-6, 5-7, 5-11 and 5-12 show diagrams of parts assembly.

5.4 Actuators and Electronics

The specifications of actuators are tabulated in Table 4. Motor controllers (MAXON EPOS-2 series) communicate through CAN bus ports. The topology of this network is a cascade connection. In order to be accommodated inside SCHEMA's body space, motor controllers are placed in the three-layered rack as shown in Figure 5-13.

Table 5.3: Mechanical parts list

Section	Unit	Sub-unit	Parts name	Drawing no.	#
HeadSection	EyeUnit		EyeUnitBaseBottom	SC_HEAD_001	1
			EyeUnitBaseFront	SC_HEAD_002	1
			EyeUnitBaseTop	SC_HEAD_003	1
			EyeUnitBaseSide	SC_HEAD_004	2
			EyeYawActuatorStayRight	SC_HEAD_005	1
			EyeYawActuatprStayLeft	SC_HEAD_006	1
			EyeYawSupport	SC_HEAD_007	2
		RightEyeUnit	EyePitchFrameRight	SC_HEAD_008	1
			EyePitchActuatorFlange	SC_HEAD_009	2
			CamSupport	SC_HEAD_010	2
			EyePitchActuatorFlangeSupportRight	SC_HEAD_011	1
		LeftEyeUnit	EyePitchFrameLeft	SC_HEAD_012	1
			EyePitchActuatorFlangeSupportLeft	SC_HEAD_013	1
		EyeLidUnit	EyeLidHarmonicStay	SC_HEAD_014	1
			EyeLidSupport	SC_HEAD_015	1
		EyeBrowUnit	EyeBrowShaft	SC_HEAD_016	2
			EyeBrowShaftSupport	SC_HEAD_017	2
			EyeBrowHarmonicFlangeRight	SC_HEAD_018	1
			EyeBrowHarmonicFlangeLeft	SC_HEAD_018-2	1
	NeckUnit		EyePitchBearingHolder	SC_HEAD_019	1
			NeckJointTop	SC_HEAD_020	1
			NeckJointBottom	SC_HEAD_021	2
			NeckJointSide	SC_HEAD_022	2
			NeckJointInner	SC_HEAD_023	1
	JawUnit		JawUnitBase	SC_HEAD_024	1
			LipPitchBaseSupportLeft	SC_HEAD_025	1
			LipSupportLeft	SC_HEAD_026	1
			LipSupportRight	SC_HEAD_027	1
			LipHarmonicFlange	SC_HEAD_028	1
	Other		NeckYawFlangeSupport	SC_HEAD_029	2
ArmSection	ShoulderUnit	ShoulderPitchUnit	ShoulderPitchStay	SC_ARM_001	2
			ShoulderBearingHolder	SC_ARM_002	4
			ShoulderSpacer	SC_ARM_003	4
			ShoulderRollStay	SC_ARM_004	2
	UpperArmUnit	RightUpperArmYawUnit	UpperArmYawStay	SC_ARM_005	2
			UpperArmYawTop	SC_ARM_006	2
			UpperArmYawSide	SC_ARM_007	4
			UpperArmYawBottom	SC_ARM_008	2
			UpperArmYawBearingHolder	SC_ARM_009	2
	ForeArmUnit	ForeArmPitchUnit	ForeArmPitchStay	SC_ARM_010	2
			ForeArmPitchBearingHolder	SC_ARM_011	2
		ForeArmYawUnit	ForeArmYawStay	SC_ARM_012	2
			ForeArmYawSide	SC_ARM_013	4
			ForeArmYawBottom	SC_ARM_014	2
	HandUnit	HandRollUnit	HandRollStay	SC_ARM_015	2
		HandUnit	HandStay	SC_ARM_016	2
			ForeArmSpacer	SC_ARM_017	4
BodySection	BoxUnit		BoxTop	SC_BODY_001	1
			BoxBottom	SC_BODY_002	1
			BoxSideBar	SC_BODY_003	4
			BoxInnerTray_Revised	SC_BODY_004	6
			PlasticTray	SC_BODY_005	2
			PlasticTraySmall	SC_BODY_006	1

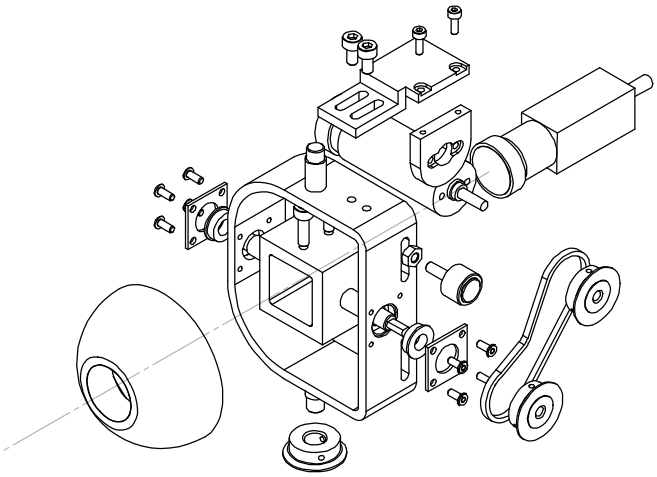


Figure 5-4: Right-eye unit assembly

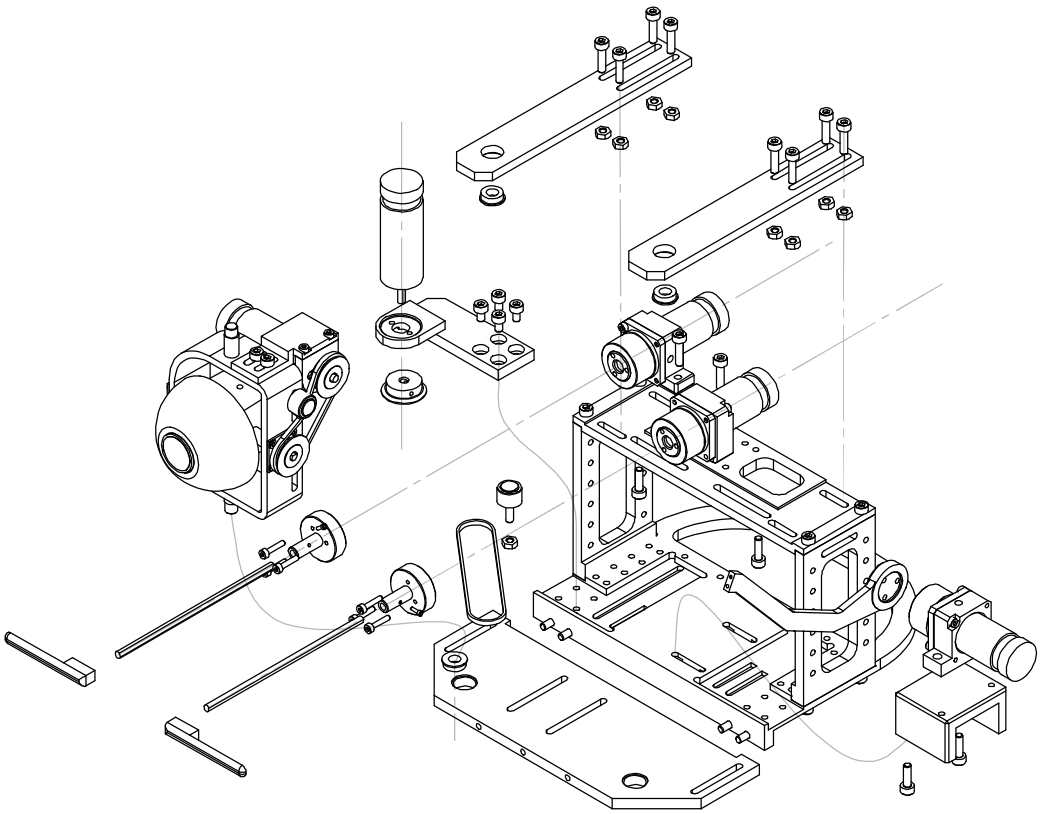


Figure 5-5: Eye-unit assembly

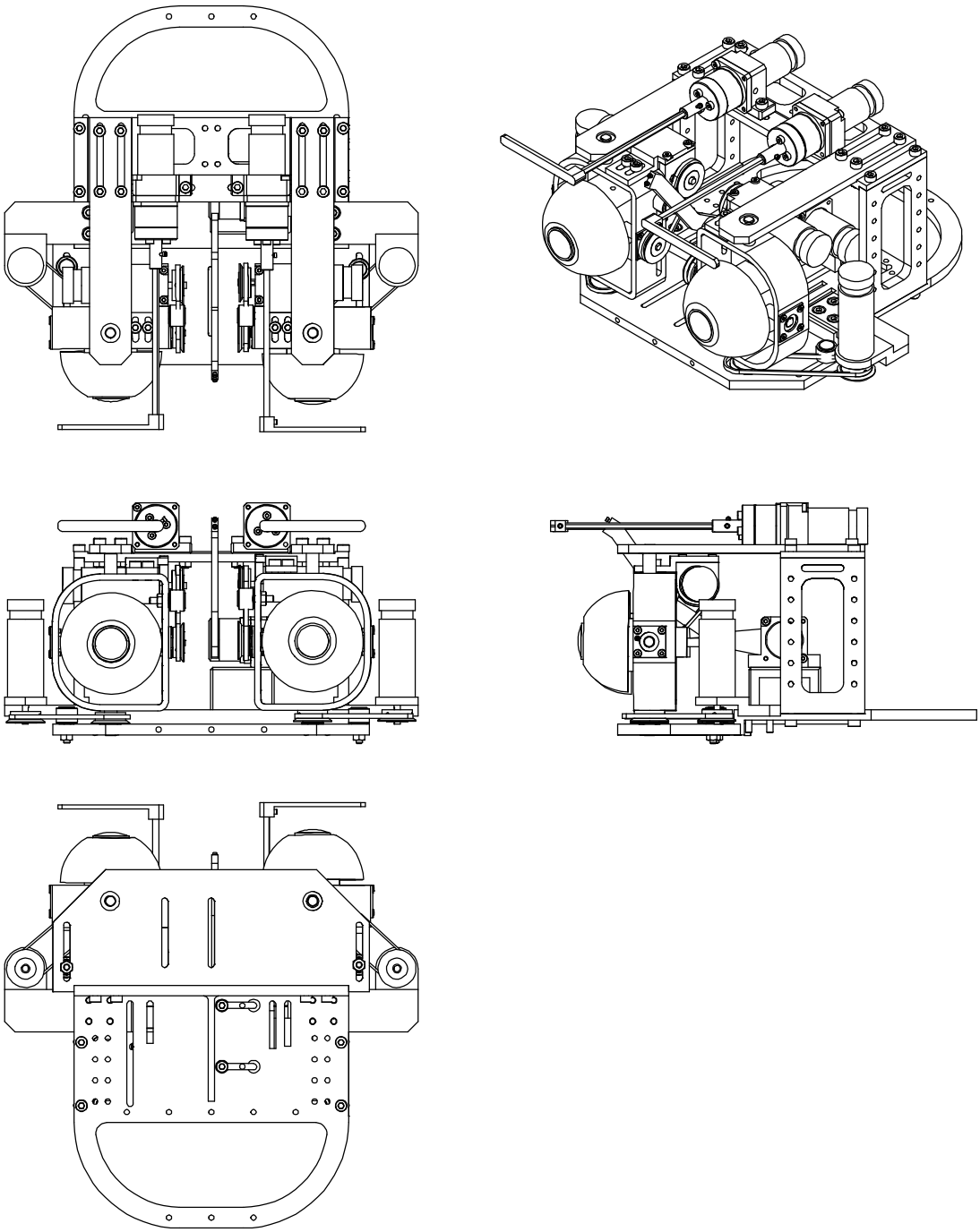


Figure 5-6: Assembled eye unit

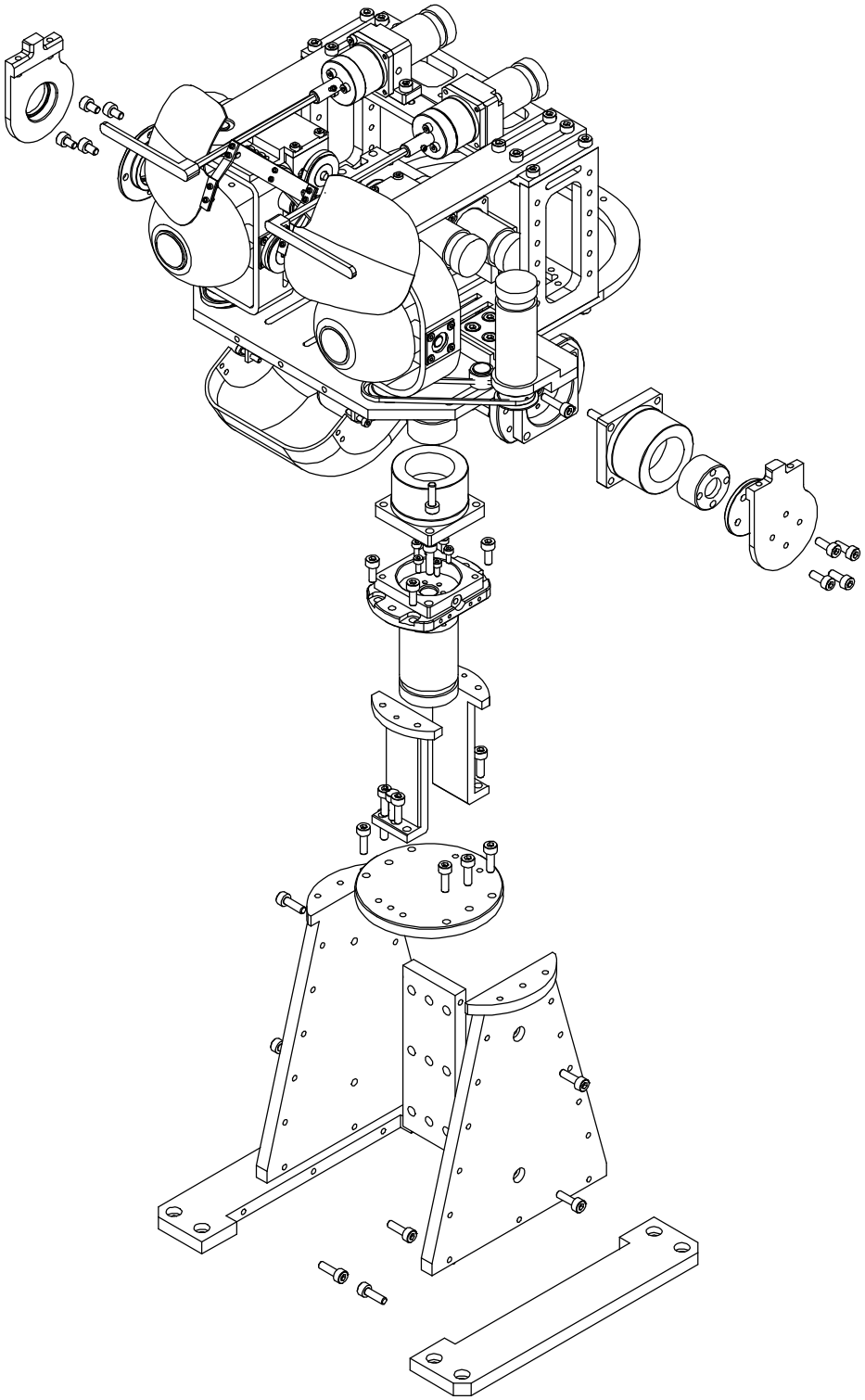


Figure 5-7: Head-unit assembly

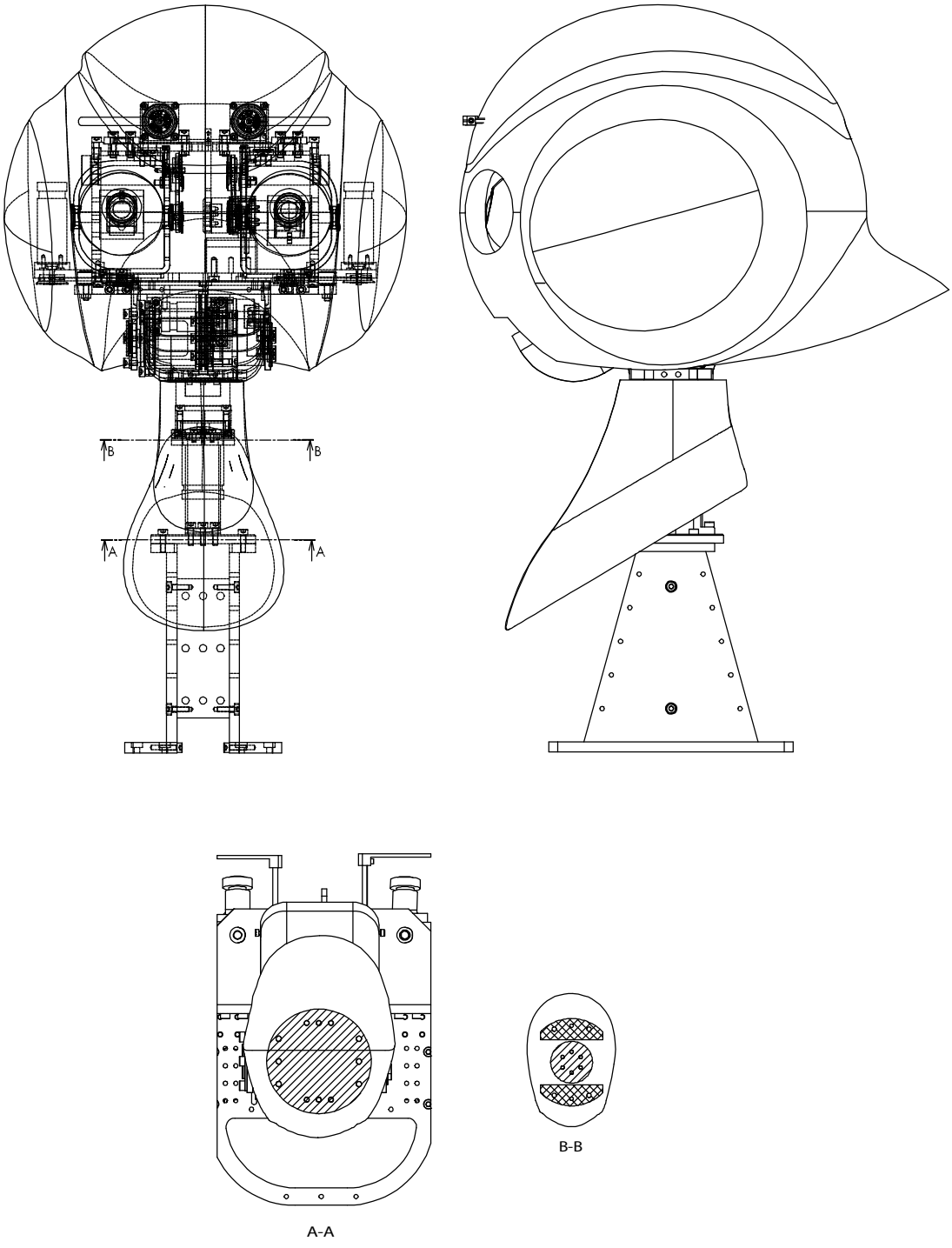


Figure 5-8: Assembled head unit with covers

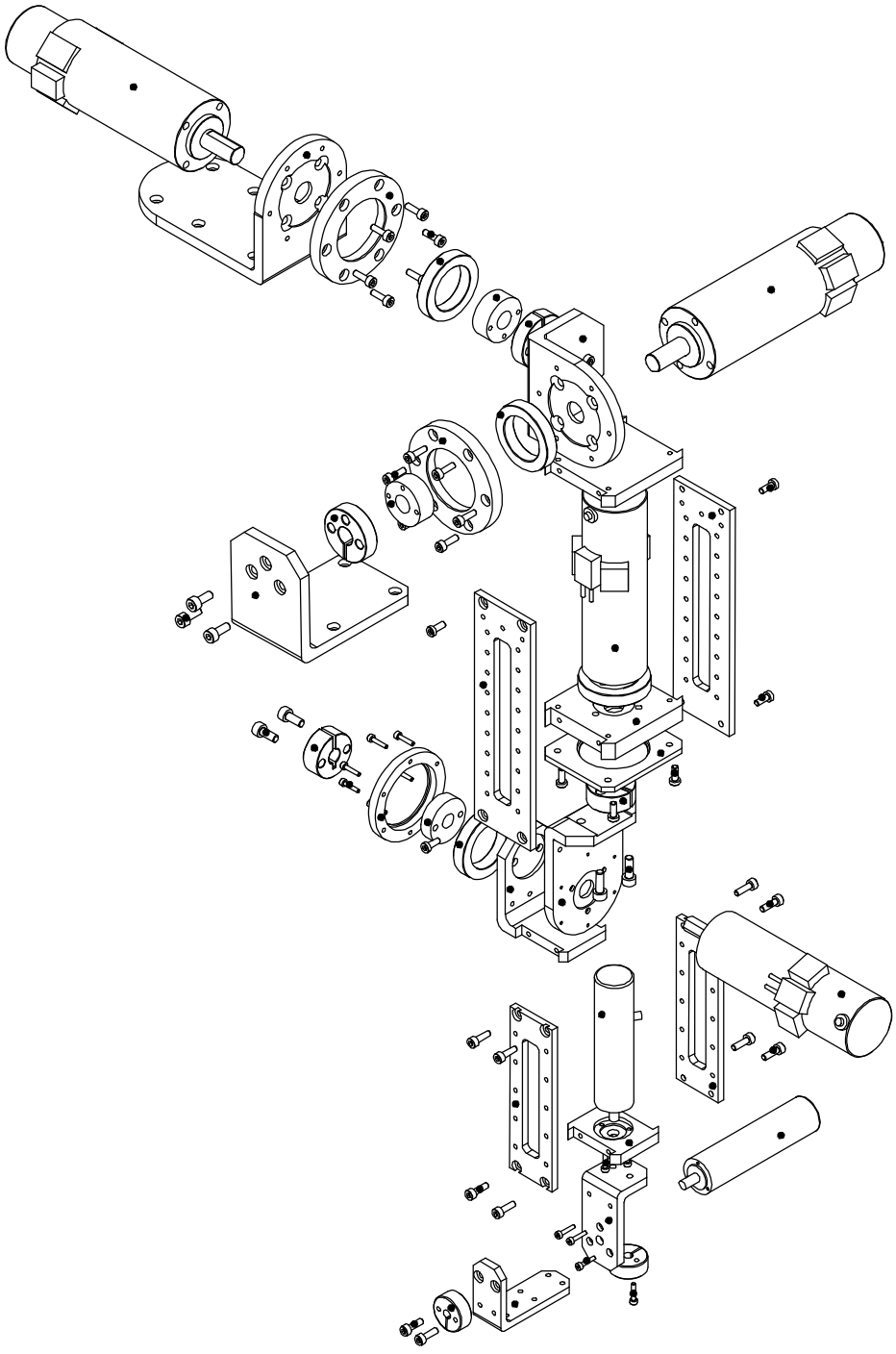


Figure 5-9: Left arm unit assembly

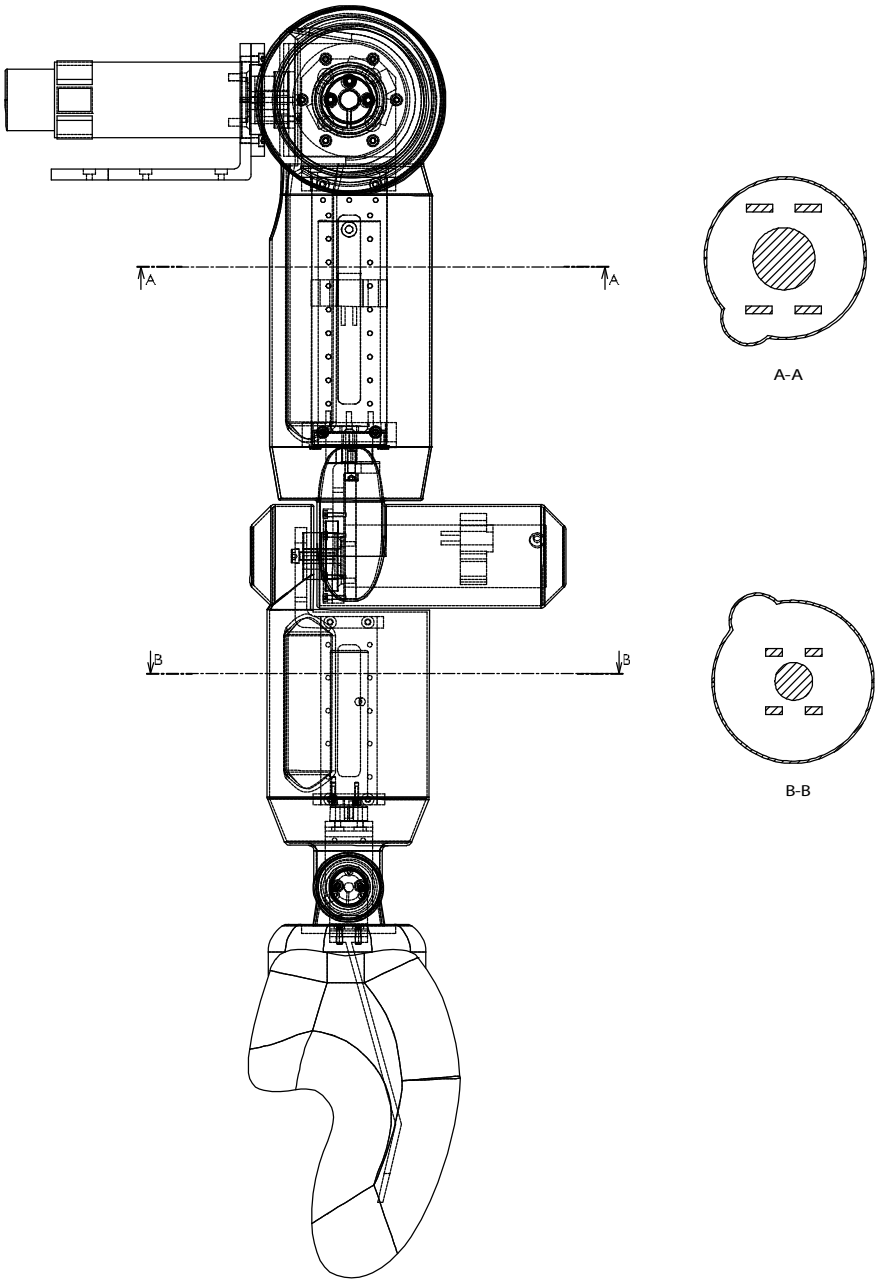


Figure 5-10: Assembled left arm unit with covers

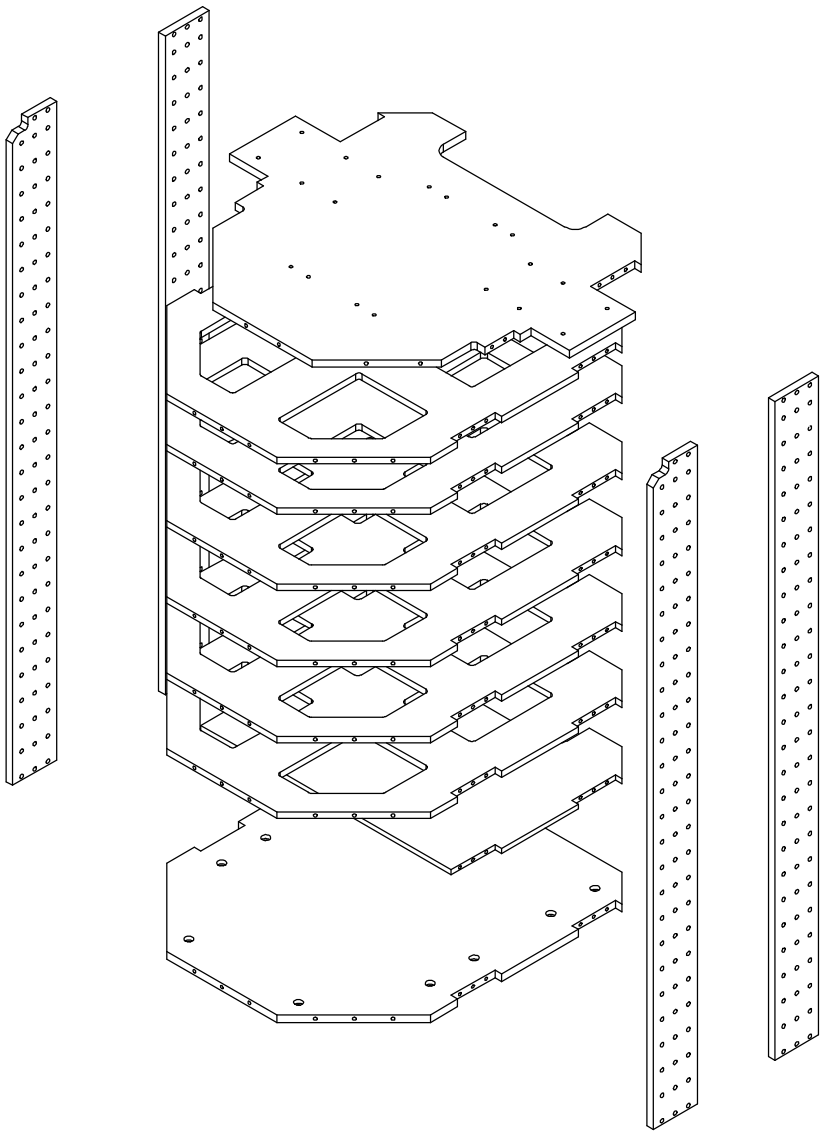


Figure 5-11: Body-unit assembly

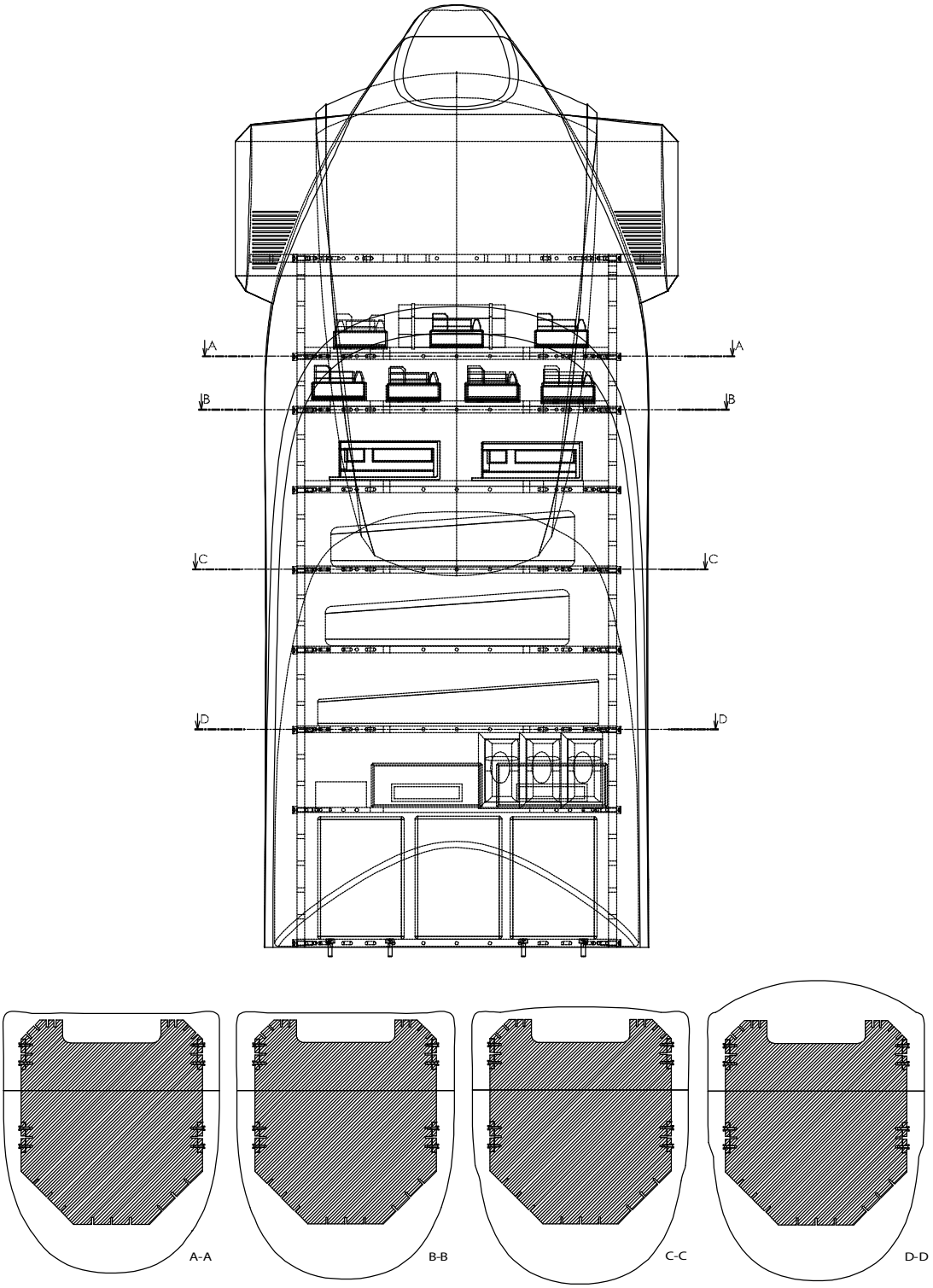


Figure 5-12: Assembled body unit with covers

Table 5.4: Specifications of actuators

DOF	Motor	Harmonic drive	Controller	Maximum velocity [rpm]	Nominal voltage[V]
Eyebrow Right	RE-max17	CSF-5-50-2XH-F-SP	EPOS 24/1	7810	24
Eyebrow Left	RE-max17	CSF-5-50-2XH-F-SP	EPOS 24/1	7810	24
Eyelid	RE-max17	CSF-5-50-2XH-F-SP	EPOS 24/1	7810	24
Eye Right Pitch	RE-max17 + GP16K	–	EPOS 24/1	7810	24
Eye Right Yaw	RE-max17 + GP16K	–	EPOS 24/1	7810	24
Eye Left Pitch	RE-max17 + GP16K	–	EPOS 24/1	7810	24
Eye Left Yaw	RE-max17 + GP16K	–	EPOS 24/1	7810	24
Mouth	RE-max17	CSF-5-50-2XH-F-SP	EPOS 24/1	7810	24
Neck Pitch	RE-max24	CSF-8-100-2XH-F-SP	EPOS 24/1	7540	24
Neck Yaw	RE-max24	CSF-8-100-2XH-F-SP	EPOS 24/1	5250	48
Right Shoulder Pitch	RH-11D-3001-E036AL	–	EPOS 24/5	3000	24
Right Shoulder Roll	RH-11D-3001-E036AL	–	EPOS 24/5	3000	24
Right Elbow Pitch	RH-8D-3006-E036AL	–	EPOS 24/1	3000	24
Right Elbow Yaw	RH-8D-3006-E036AL	–	EPOS 24/1	3000	24
Right Wrist Yaw	RH-5A-5502-E036AL	–	EPOS 24/1	4500	12
Right Wrist Roll	RH-5A-5502-E036AL	–	EPOS 24/1	4500	12
Left Shoulder Pitch	RH-11D-3001-E036AL	–	EPOS 24/5	3000	24
Left Shoulder Roll	RH-11D-3001-E036AL	–	EPOS 24/5	3000	24
Left Elbow Pitch	RH-8D-3006-E036AL	–	EPOS 24/1	3000	24
Left Elbow Yaw	RH-8D-3006-E036AL	–	EPOS 24/1	3000	24
Left Wrist Yaw	RH-5A-5502-E036AL	–	EPOS 24/1	4500	12
Left Wrist Roll	RH-5A-5502-E036AL	–	EPOS 24/1	4500	12

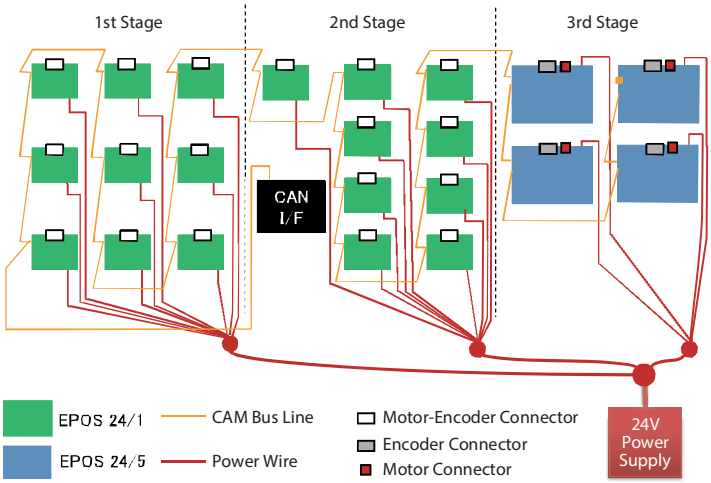


Figure 5-13: Electronics setting

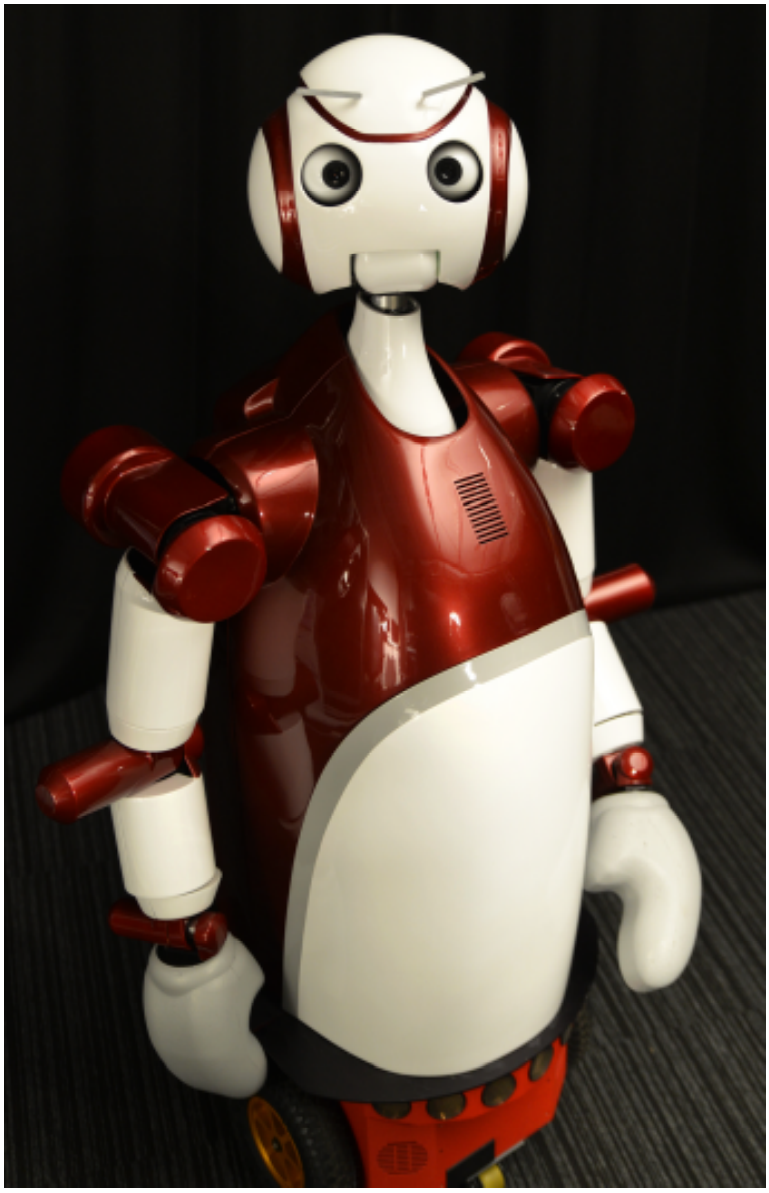


Figure 5-14: SCHEMA: Multiparty conversation oriented robotic platform

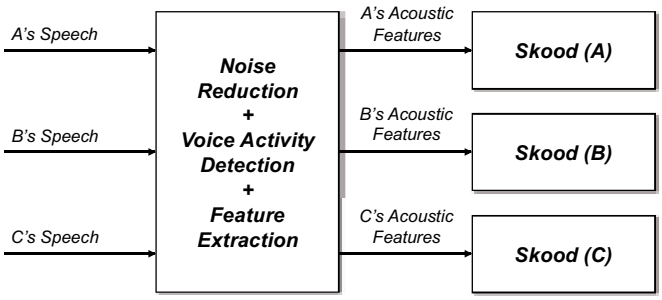


Figure 5-15: Speech recognition software module setting

5.5 Sensor and Motor Modules

5.5.1 Speech Recognition

Device Setting

In our experimental setting, a conversational group was organized with three participants excluding the robot. We used separate speech recognizers for each participant. The system also records ambient noises with speech sources for noise reduction. Each participant wears a nondirectional wireless microphone. A nondirectional wired microphone is used for ambient noises. After each speech, sounds from the sources are amplified and converted from analog to digital form, and provided as inputs to a PC.

Speech Recognition Software Module Setting

After the speech from sources are recorded using the devices in the manner described above, they are processed by the speech recognition system consisting of two major modules, namely, the gmfccserver and the skood. The gmfccserver is a Mel-frequency cepstral coefficient (MFCC) feature extractor, which is locally networked with the skood, a speech recognizer. Each time the gmfccserver extracts features, the skood decodes as a recognition result. The result is sent to other modules though MONEA. Figure 5-15 shows the speech recognition software module setting.

5.5.2 Action Player

As a behavior realizer, we implemented SCHEMA Action Player that synchronizes motor control and speech synthesis modules. The motor control module is implemented along a layered messaging model consisting of a hardware layer, controller layer, DOF layer, pattern layer, and action layer.

1. Hardware Layer

This layer corresponds to hardware devices. Switches embedded on the robot enable switching the electric power supply on or off.

2. Controller Layer

This layer corresponds to the SDKs of EPOS controllers and the turret. When electric current is supplied from the hardware layer to the EPOS controllers, they are disable to control. When they are initialized to be enabled, the following commands are available in this layer.

- Initialize to the original point
- Getting the current angle [qc]
- Getting the current velocity [rpm]
- Controlling the DoF [qc]
- Controlling the DoF (angle [qc], velocity [rpm])
- Controlling the DoF (angle [qc], velocity [rpm], acceleration [rpm/s], deceleration [rpm/s])
- Stopping control

During runtime, if a certain defect is detected, the system goes into the fault state, and the EPOS controller is disable to control.

3. DOF Layer

This layer corresponds to DOF control. When the system is enabled, the following commands are available in this layer.

- Getting the current angle of a DOF [°]
- Getting the current velocity of a DOF [° /s]
- Controlling a DOF (angle [qc])
- Controlling a DOF (angle [qc], velocity [rpm])
- Controlling a DOF (angle [qc], velocity [rpm], acceleration [rpm/s], deceleration [rpm/s])
- Stopping control of a DOF

When the second layer (controller layer) is disabled, this layer is also disabled.

4. Pattern Layer

This layer corresponds to playing patterns. Patterns consist of absolute angle patterns and relative angle patterns. This layer does not resolve conflicts of patterns, but only plays along given commands. If it receives multiple patterns simultaneously, patterns are synthesized and played. Although it is possible to cancel playing of patterns, an angle of stopping position is not secured. When the system is enabled, the following commands are available in this layer.

- Getting the current angle [°]
- Getting the current velocity [° /s]

- Playing a pattern
- Canceling a pattern

When the third layer (DOF layer) is disabled, this layer is also disabled.

5. Action Layer

This layer corresponds to generating patterns that are passed to the pattern layer. The actions are of the following two types:

- Periodic Action
  Behaviors involving fine motor coordinations such as nodding. Periodic actions are repeated either eternally or for a certain duration.
- Non-periodic Action
  Behaviors without repetitions such as raising a hand and looking at an object.

When the system is enabled, the following commands are available in this layer.

- Checking the feasibility of a certain action
- Playing action
- Canceling a certain playing action
- Canceling all actions

When the fourth layer (pattern layer) is disabled, this layer is also disabled.

5.5.3 Turret Control

When the turret is enabled, the current position is set as the origin point. We define the coordinate system of the turret as shown in Figure 5-16. Here, the distance is expressed in mm, and rotation is expressed in deg. A clockwise rotation is considered positive.

Commands of turret control

- **GoTo(x, y, vel, acc, dec)**
  - Moving to a point (x, y) with vel [mm/sec], acc[mm/sec/sec], dec[mm/sec/sec], in an arbitrary angle
  - vel dec can be abbreviated
- **GoToForward(x, y, vel, acc, dec)**
  - Moving to a point (x, y) with vel [mm/sec], acc[mm/sec/sec], dec[mm/sec/sec], in a frontal angle
  - vvel dec can be abbreviated

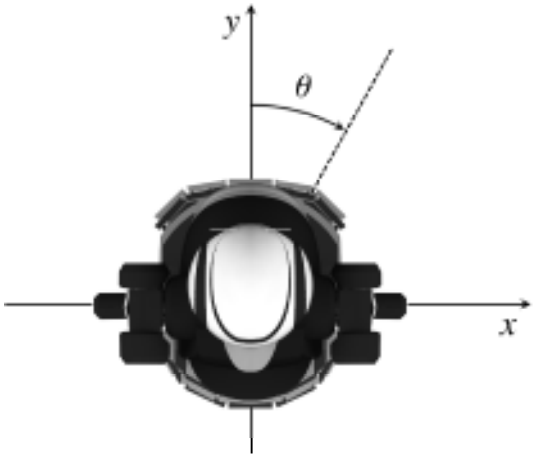


Figure 5-16: Coordinate system of the turret (top view).

- **GoToBackward(x, y, vel, acc, dec)**
  - Moving to a point (x, y) with vel [mm/sec], acc[mm/sec/sec], dec[mm/sec/sec], in a back angle.
  - vvel dec can be abbreviated
- **Head(θ, θ')**
  - Rotating to an angle θ with gel θ'[deg/sec]
  - θ' can be abbreviated
- **GoToAndHead(x, y, θ ,vel, acc, dec, θ')**
  - Moving to a point (x, y) with vel[mm/sec], acc[mm/sec/sec], dec[mm/sec/sec], and rotating to an angle θ with gel θ' [deg/sec]
  - vel θ' can be abbreviated

5.5.4 Speech Synthesis

TOSHIBA corporation provided a speech synthesizer called ToSpeak that is customized for our conversational system. In this section, we describe the method of connecting the ToSpeak synthesizer and other modules though MONEA. Figure 5-17 shows the overview of the connection. When the speech synthesis program, named the sch_monea_synthesizer, gets a request from other modules though MONEA, it generates the speech waveform and a duration of phonemes, and then sends them to the Action Player. The Action Player plays a lip-sync action based on the duration of phonemes. The sound is emitted from a speaker embedded on SCHEMA's chest.

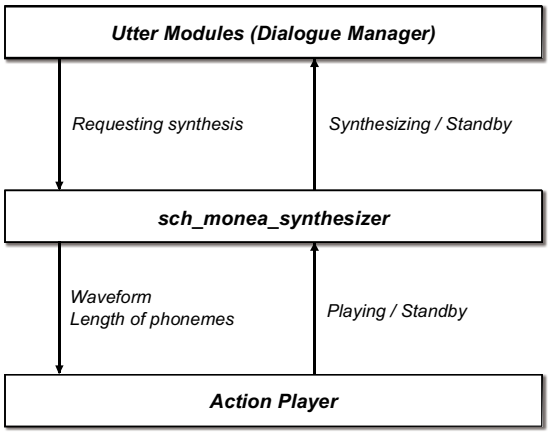


Figure 5-17: Interfaces of the speech synthesis module.

5.6 Network Middleware

Network middleware abstractions are defined in terms of protocols in order to organize sophisticated distributed systems. Most software for robotic platforms are developed on top of middleware packages that support modularity and hardware interfacing. They also generally ensure independence from the types of operating systems and development tools used. In this section, we review some popular middleware packages namely ROS, YARP, VHMsg, and MONEA used in robotics and agent research communities.

5.6.1 Existing Popular Middlewares: ROS, YARP, VHMsg

ROS (Robot Operating System)[1] is an open-source network middleware for general purpose robotic systems. It defines hardware abstraction, low-level device control, implementation of common functionality, message passing between processes, and provide a package management system. It employs a peer-to-peer networking model. Processes can be grouped into packages and stacks, which can be easily shared and distributed. For purposes of sharing and collaboration, ROS is designed to be as thin as possible so that codes can be reused in other robot software frameworks. ROS supports many popular programming languages, including C++, Python, Lisp, Java, and Lua, to write modules ROS also supports code repositories, and provides a built-in unit/integration test framework called rostest. Such development environments and an ecosystem make ROS a very popular middleware for many robots worldwide.

YARP (Yet Another Robot Platform) is mainly used for the iCub developed by Italian Institute of Technology[2]. YARP is a thin library like ROS, which allows multi-platforms and multiple IDEs. The whole design of iCub hardware (drawings, schematics, specifications) and its software (both middleware and controllers) is distributed according to the GPL or the LGPL licenses. Software interfaces between modules are defined in terms of YARP ports and the type of data to receive/send. While YARP itself is written in C++, many popular program language can be allowed to develop modules, including Python, Java, Ruby, C#.

VHMsg (Virtual Human Toolkit Message) is for the Virtual Human Toolkit developed by the University of Southern California Institute of Creative Technologies (ICT), which defines a protocol and provides an

[1] http://www.ros.org/
[2] http://wiki.icub.org/yarp/

API wrapper built on top of ActiveMQ[3]. The Virtual Human Toolkit is a framework of computer graphics based embodied conversational agents. Messages are broadcasted as default, Like ROS and YARP, VHMsg allows many popular programing languages to develop new modules. A standard set of message types is used by existing modules, and it is very easy to create new message types (e.g. vrSpeech, vrExpress).

5.6.2 MONEA: Message-Oriented NEtworked-robot Architecture

As the network middleware package for the SCHEMA platform, we employ MONEA (Nakano et al., 2006) developed by our group (Perceptual Computing Group, Waseda University). Like other middlewares for multifunctional robots, MONEA was designed as a thin middleware, supporting the "bazaar-style" development model. MONEA was implemented as an information sharing framework, named "networked-whiteboard" model, along with a message passing framework via a peer-to-peer virtual network. MONEA architecture provides the asynchronous message passing and information sharing framework for robot system construction using networked resources. The major concepts of MONEA are described as follows.

- **Multicast: Networked-Whiteboard Model**
  Intermodule information sharing is realized as this model. Each module has its "Whiteboard," writable area to write its internal state and generated information. The whiteboard allows both disclosed and undisclosed areas to control transparency of area in a group. The networked-whiteboard model has two major roles: publisher and subscriber, which have the following operations to guarantee information atomicity.

  - Publisher
    1. *Write*: writing data on the Whiteboard
    2. *Commit*: confirming a set of written data (guarantees atomicity of writing data)
  - Subscriber
    1. *Update*: obtaining the latest updated data (guarantees atomicity of reading data)
    2. *Read*: reading data from the Whiteboard

  This model allows two different modes defined in terms of message sending timing: *PUSH* and *PULL* Mode. In the *PUSH* Mode, publisher sends messages as soon as it is committed. In the *PULL* Mode, publisher sends messages when it gets a request message from a subscriber. A developer can choose one of the two modes in terms of the frequency of commit and update in a system.

- **Unicast : Processing Request Model**
  In this mode, each module can synchronously send a processing request message to other modules to perform certain processing. While the actual performance of a sender's requests depends on the receiver's state and message contents, the sender module can know the performing status by observing the receiver's state.

- **Interest-Oriented Module Group Model**
  In order to reduce complexity, module developers can define interest-oriented module groups that can be classified in terms of common interests (e.g., vision, speech). Within a group, modules' roles, disclosing information, and available processing requests can be defined.

[3] http://activemq.apache.org/

Module developers create a module definition file as an XML format for each module. This file contains a list of disclosable properties, processing request to be handled, groups and disclosures for each group, and remote modules. The definition file sample for the publisher is as follows.

```
<?xml version="1.0"?>
   <moduleContext>
   <local>
      <property name="status"/>
      <property name="expression"/>
      <property name="direction.x"/>
      <property name="direction.y"/>
      <property name="velocity"/>
      <method name="exec">
      <param name="id"/>
   </method>
   </local>
   <disclosure>
   <group>RobotControl</group>
   <role>SCHEMA</role>
   <propertyRef name="status"/>
   <propertyRef name="expression"/>
</disclosure> </moduleContext>
```

The definition file sample for the subscriber is as follows.

```
<?xml version="1.0"?>
   <moduleContext>
   <local>
      <property name = “id"/>
      <property name = “name"/>
      <property name="status">
      <description> run or x</description>
      </property>
   </local>
   <disclosure>
   <group>Dialogue</group>
   <role>main</role>
   <propertyRef name="status"/>
   </disclosure>
   <remote name= “robot ” >
   <group>RobotControl</group>
   <role>SCHEMA</role>
    </remote>
</moduleContext>
```

An actual definition file used for our experiment is as follows.

```
<?xml version="1.0"?>
<moduleContext>
    <conf>
```

```
        <moduleName>main</moduleName>
        <description>mainModule</description>
    </conf>
    <remote name="synthesizer">
        <group>speechSynthesis</group>
        <role>speechSynthesizer</role>
    </remote>
    <remote name="action">
        <group>robot</group>
        <role>actionPlayer</role>
    </remote>
    <remote name="imageSensor">
        <group>camera</group>
        <role>imageSensor</role>
    </remote>
    <remote name="kinectSensor">
        <group>kinect</group>
        <role>kinect_participant_A</role>
    </remote>
    <remote name="kinectSensor">
        <group>kinect</group>
        <role>kinect_participant_B</role>
    </remote>
    <remote name="kinectSensor">
        <group>kinect</group>
        <role>kinect_participant_C</role>
    </remote>
    <remote name="speechRecognizer_A">
        <group>speech</group>
        <role>recognizer_participant_A</role>
    </remote>
    <remote name="speechRecognizer_B">
        <group>speech</group>
        <role>recognizer_participant_B</role>
    </remote>
    <remote name="speechRecognizer_C">
        <group>speech</group>
        <role>recognizer_participant_C</role>
    </remote>
    <remote name="POMDPModel">
        <group>pomdp</group>
        <role>actionSelection</role>
    </remote>
    <local>
        <property name="observation">
            <group>pomdp</group>
            <description>observation</description>
            <value>null</value>
        </property>
```

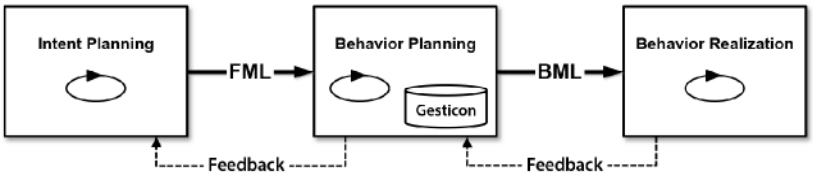


Figure 5-18: SAIBA framework for multimodal behavior generation

```
        <property name="observationId">
            <group>pomdp</group>
            <description>observationId</description>
            <value>0</value>
        </property>
    </local>
    <disclosure>
        <group>pomdp</group>
        <role>model</role>
            <propertyRef name="observation"/>
            <propertyRef name="observationId"/>
    </disclosure>
</moduleContext>
```

5.7 Discussions on Higher Level Protocols

5.7.1 SAIBA : Multimodal Behavior Generation Framework

While a network middleware has responsibilities of secure massaging, a conversational system also requires higher general protocols to manage time-critical understanding and production process with high flexibility. One actively discussed higher level protocols for ECAs is the SAIBA (Situation, Agent, Intention, Behavior, Animation) framework, which aims at unifying a multimodal behavior generation framework for ECAs[4] (Kopp et al., 2006; Vilhjálmsson et al., 2007). The SAIBA framework specifies multimodal behavior generation, consisting of processing stages on three different levels: (1) planning of a communicative intent, (2) planning of a multimodal realization of this intent, and (3) realization of the planned behaviors. Figure 5-18 depicts the overview of the SAIBA framework.

The behavior markup language (BML) is an interface between behavior planning and behavior realization. It provides a general description of multimodal behaviors for embodied agents. The core of the BML standard defines the form and use of BML blocks, synchronization, feedback about the processing results of BML messages, and a number of generic basic behaviors. A BML realizer is responsible for executing a multimodal plan incrementally scheduled. Welbergen et al. proposed the AsapRealizer 2.0, a dynamic BML behavior realizer that has several fluent behavior realization capabilities. It allows incremental multimodal utterance construction, including speech, gaze, and facial expression (van Welbergen et al., 2014).

The functional markup language (FML) is an interface between intent planning and behavior planning describes communicative and expressive intent without any reference to physical behavior using FML. It

[4]http://www.mindmakers.org/projects/saiba/wiki

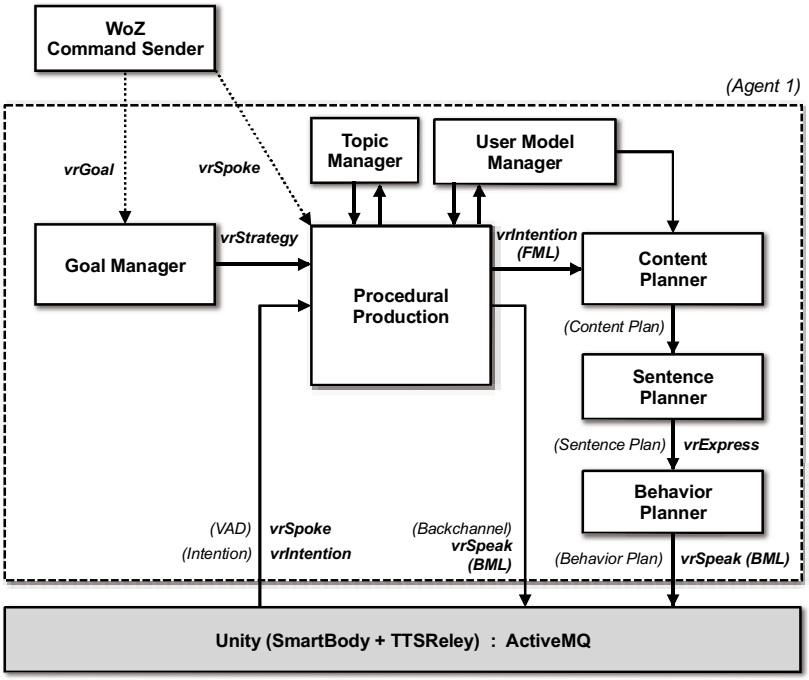


Figure 5-19: Architecture of multi-agent simulator

is meant to provide a semantic description that accounts for the aspects that are relevant and influential in the planning of verbal and nonverbal behavior (Lee et al., 2008; Heylen et al., 2008; Cafaro et al., 2014). There has been an ongoing discussion about FML in a series of targeted workshops, including contextual information and personal characteristics, communicative actions (e.g. dialogue acts, grounding actions, and turn taking), and emotional and mental states. In order to extend the SAIBA framework to the multiparty interaction model, some other issues of FML are also being discussed (e.g. participation structure).

5.7.2 Multi-Agent Simulator

Ravenet et al. presented preliminary implementation of a multiparty agent simulation system allowing agents to exhibit a variety of nonverbal behaviors (e.g., gestures, facial expressions, proxemics), depending on the interpersonal attitudes in a group conversation(Ravenet et al., 2014). The model is based on a combination of a theoretical framework and a corpus. Inspired by these works and the design guideline of the SAIBA Framework, we developed a prototype of a conversational multi-agent simulator for both dyadic and multiparty situations. Figure 5-19 shows the architecture of the simulation system. In order to avoid recognition (e.g., intention, speech recognition, facial recognition) errors, outputs of an agent will be directly sent to other agents as sensory information. All modules are implemented in line with the VHMsg protocol. The procedural production module produces *vrIntention* as a form of BML, defining *agent_id*, *task_goal*, *group_maintenance_goal*, and *facilitation_strategy dialog_act*, sending it to the content planner. The content planner refers the user model to obtain a user's profile and interests, and generate a certain logical form of content plan. The sentence planner generates a sentence plan sent to the behavior planner.

In the current version, we have used the jSPaRKy v2.0 sentence planner developed by AT&T [5] for sentence planning, and he nonverbal behavior generator (NVBG) developed by the USC ICT [6] (Hartholt et al., 2013), for behavior planning. The agents are animated on the ICT's SmartBody [7] connected to the Unity 3D game engine [8].

5.8 Conclusions and Future Work

In this chapter, as a robotic platform that can facilitate multiparty conversation, we presented the design of the SCHEMA robotic platform, including both hardware and software architectures. We presented exterior, mechanical, and software designs of the SCHEMA robot. We also extensively discussed higher level protocols of conversational robot systems. Based on the SAIBA guideline, we implemented a multi-agent simulator. Future work will include extending the SAIBA framework towards multiparty interaction, and applying the extended framework to an implementation of the SCHEMA system.

[5] http://www.research.att.com/archive/people/Stent_Amanda_J/library/documents/sparky2.0/index.html
[6] https://confluence.ict.usc.edu/display/VHTK/NonVerbal+Behavior+Generator
[7] http://smartbody.ict.usc.edu/
[8] http://unity.com/

"Palliative care is an approach that improves the quality of life of patients and their families facing the problem associated with life-threatening illness, through the prevention and relief of suffering by means of early identification and impeccable assessment and treatment of pain and other problems, physical, psychosocial and spiritual."

The World Health Organization (WHO),
"*Definition of Palliative Care*"

6 Applications

We propose a robot that promotes an enjoyable party game as a facilitation robot system. In this task, a robot participates in a quiz game as one of participants and tries to promote the other participants' enjoyment of the game. The functions implemented in the robot are as follows. (1) The robot participates in the group's communication using its basic group conversation functions. (2) The robot performs the game according to the rules of the game. (3) The robot facilitates communication using its proper actions, depending on the game's and participants' situations. We conducted a real field experiment in which a prototype system participates in a quiz game with elderly people in an elderly day-care center. The robot successfully entertained the people with its one-hour demonstration. We also evaluated its interactions with subjects in the Nandoku quiz game using video analysis and a semantic differential (SD) method that utilizes questionnaires. The results of the SD method indicate that the subjects were more pleased and felt the game was noisier when the robot participated. The results of the video analysis indicate that the smiling duration ratio is greater with the robot's participation. These results evidence the robot's communication activation function in the party game.

6.1 Introduction

We propose the participation of a robot in the group communication and activation of a game. A substantial amount of daily communication involves group communication, in which multiple people participate. Therefore, a robot should be able to naturally participate in group communication when completing given tasks. In this study, we focus on a quiz game, which is a recreation provided at a day-care center for the elderly, and develop a robot that can participate in the game and activate participation in others.
As we reviewed in chapter 2, conventional robots and spoken dialogue systems have been assumed to be dyadic (two-party) communication. This assumption is so well established that developed systems cannot be directly applied to group communication. There are several applications of multiparty conversation robot systems, such as meeting support, learning tutors, party games, and elderly care. Meeting support appli-

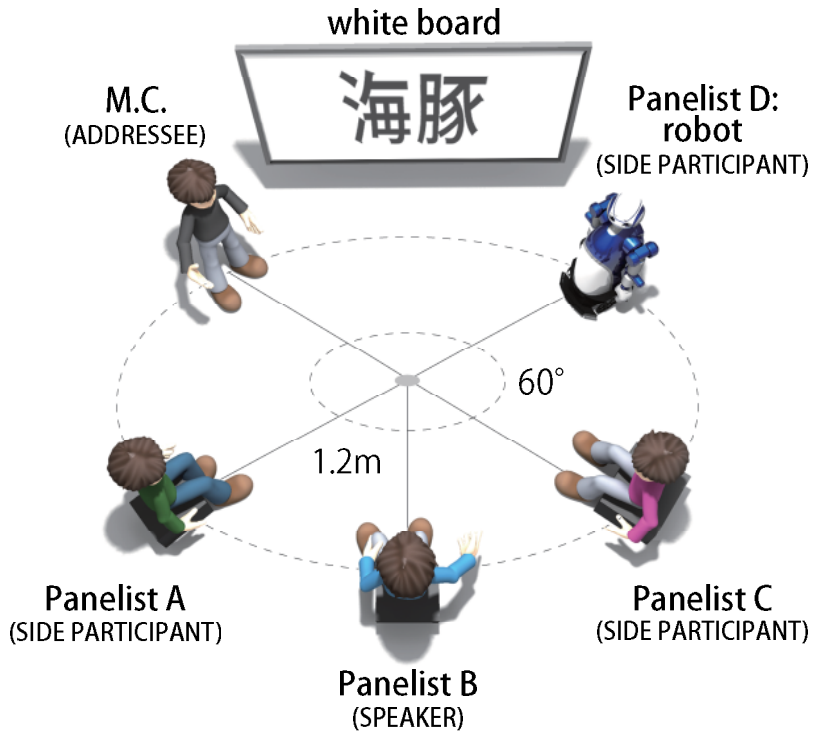


Figure 6-1: Illustration of the Nandoku game setting in which each participant has a role (speaker, addressee, or side participant) and a robot participates in the game as one of panelists in order to directly and indirectly facilitate the game

cations have been considered in the context of computer-supported cooperative work (CSCW) (Dubs and Hayne, 1992), group support systems (GSSs) (Bostrom et al., 1993), and group decision support systems (GDSSs) (Watson et al., 1988). Typical games involve the gathering together and enjoyment of several people, such as cards, Sugoroku, or quiz games. These kinds of games are called party games. In our study, we focus on the development of a robot that participates in a party game and initiates communication among the participants. Many studies have been performed on robots used for entertainment and communication. Ifbot is a robot that can talk with people, but it is assumed to be a one-to-one conversation model (Kato et al., 2004). QRIO is a small biped entertainment robot that displays some sort of performance, such as dancing (Ishida, 2004). Paro is a seal-like robot that creates a relaxing environment by initiating communication among elderly people in a method similar to that of an animal (Wada and Shibata, 2006). Experimentation has been performed on Robovie in fields, such as elementary schools. Although it affects human relationships in long-term experimentation, short-term communication, such as a game, has not been discussed (Kanda et al., 2007). PaPeRo is a robot that has various short-term communication abilities. It can participate in a game, such as a quiz, but its communication activation potential has not yet been discussed (Osada et al., 2006).

In this study, we focus on a quiz game, which has one master of ceremony (MC) and several panelists, and develop a robot that participates as one of panelists. This game provides recreation in an elderly day-care center and is expected to initiate communication. In order to make a game effective, it needs a

very good MC. Otherwise, the participants will not be engaged and may not be motivated to answer the quiz questions. Thus, we develop a robot that participates in the game as one of the panelists, just like the elderly participants, and activates the game. In addition to participation in the group's communication, the robot needs another ability in order to involve the other panelists and motivate them to appropriately participate in the game. We aim to initiate communication, and mechanisms for implementing this are also investigated. In this chapter, we propose a robot architecture that comprehensively addresses these principles of communication. Using this system, we conduct a field experiment in an elderly day-care center. We also conduct an experiment in order to evaluate this communication activation system's effectiveness and to analyze its violations, which decrease the active communication. Kanda et al. evaluated how a robot's eye gaze affects subjects' impressions using a semantic differential (SD) method (Kanda et al., 2001). Based on prior methodology, we subjectively and objectively evaluate effectiveness using the SD method and video analysis.

6.2 *Nandoku*: Elderly Care Application

6.2.1 Robot as Communication Activator

In this chapter, we present the *Nandoku* quiz game implemented on a conversational robot. Kanji characters are typically used in Japanese literature. The same Kanji character reads differently when it constructs a word with other Kanji characters. In general, people know the pronunciation and meaning of many words but often do not know how to write using Kanji characters. Thus, many Japanese speakers find Nandoku both difficult and interesting.

In order for a robot to participate in a party game and initiate communication between the other participants, the robot must satisfy the constraints described in the previous section and select the action that best activates communication. We adopt a strategy that requires the robot to select an action that activates communication. In this way, the robot is able to naturally initiate communication in the game as a participant. The robot is one of the panelists, as shown in Figure6-1.

6.2.2 Group Communication Constraints

The robot must play the role of *speaker*, *addressee*, and *side participant* at any given time during the game. The action it selects is subjected to the following constraints for each role.

- *speaker*
  - replying to the former speaker (*essential*)
  - looking at and/or pointing to target objects in order to demonstrate attention (*essential*)
  - looking at addressee (*essential*)
- *addressee*
  - nodding in response to speaker's utterance (*essential*)
  - looking at speaker or target objects (*essential*)
- *side participant*
  - looking at target objects (*essential*)

  - nodding in response to speaker's utterance (*optional*)
  - looking at next speaker (*optional*)

The *essential* constraints are strong ones that the robot must always satisfy and the *optional* constraints are weak ones that the robot should satisfy when it can.

6.2.3 Task Constraints

The proposed robot participates in a quiz game. In this type of game, there is one MC and several panelists. In this study, the robot acts as one of the panelists. The constraints that the robot should satisfy as a panelist are as follows.

- It should answer the questions presented by the MC.
- It should answer questions when it possesses appropriately prepared answers.
- It should make an effort to prepare an answer when the question is difficult. For example, it should ask the MC for a hint.

Note that these actions should be done within the satisfactory constraints of group communication. For example, the robot must not disturb another participant's response by answering the question itself.

6.2.4 Communication Activation Constraints

In general, game participants attempt to win the game. In a quiz game, this means that the participants answer a question as soon as they have an answer. The aim of this study, however, is not to create a robot that wins the game but one that initiates communication.

In a quiz game, an activated situation is one in which the panelists actively and frequently answer the questions. The panelists should also sincerely attempt to answer the questions.

In order to initiate communication, the robot should select an action that creates multiple opportunities for other panelists to answer by:

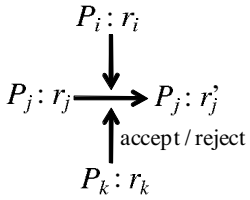
- encouraging another participant to answer,
- asking for a hint from the MC, who knows the correct answer,
- saying something that is itself a hint, or
- giving close answers that are not correct.

The reason we would develop a robot that gives close but not correct answers is because the quiz finishes when someone answers correctly.

6.2.5 Request-Answer Model

We propose an effectiveness model for the participants' behaviors when a robot is included in the participation structure. Some participant behaviors possess functions that change the states of other participants. For instance, when a speaker gazes at a side participant, he or she has the ability to change the role of the side participant into an addressee. We use the phrase "ability to change" because the speaker needs the side participant's acceptance in order to change the role. These requests that the participants present to

one another, along with their acceptances and rejections, change the roles of the other participants and are repeated throughout the game. Therefore, we propose the description methodology shown in Figure 6-2.



$P_{i\cdots k}$: participants ($i \neq k$)
$r_{i\cdots k}$: role of each participant

Figure 6-2: Function of participation roles

Requests are categorized in two ways: self-assignment (Which role do I want to be?) and assignment to other participants (Which other participants' roles do I want to change?). For instance, the behavior of maintaining a turn is the speaker's request to assign the role of speaker to him or herself(Figure 6-3(a)). An addressing behavior can be regarded as the speaker's request to assign the addressee a new role of side participant(Figure 6-3(b)).

Responses can be categorized as acceptances or rejections. A behavior-accepting turn could be the addressee's acceptance of the speaker's request to assign the speaker the role of addressee(Figure 6-4(a)). On the other hand, an example of a behavior-rejecting turn is an addressee's rejection of the speaker's request to assign the speaker the role of addressee(Figure 6-4(b)).

6.2.6 Functions of Behaviors in Quiz Game Task

In Nandoku, panelists not only answer questions but also encourage other panelists to answer. Therefore, the functions of the robot's behavior are determined with the goal of progressing the game, and we define the following four behaviors.

1. `ANSWER`: function to offer answer
2. `ASK_HINT`: function to ask MC for hint
3. `LET_ANSWER`: function to encourage other panelists to answer
4. `INFORM`: function to offer trivia information depending on the question

6.2.7 Function of Behaviors in Communication Activation

In Nandoku, communication is activated by giving other panelists the opportunity to answer questions. This successful situation can be realized not only by directly encouraging someone to answer but also by giving hints to other panelists or asking the MC for a hint. Communication can also be achieved when the robot reacts to an MC's statement, attracting the attention of other panelists.

Additionally, when either the MC or robot offers interesting information, the situation should be activated. However, because this function should be included in the functions that progress the game, we will incorporate it in that discussion.

We define the following four communication-activating functions:

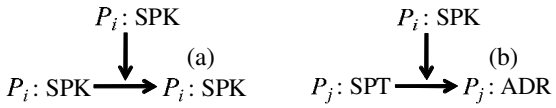


Figure 6-3: Example of request: (a) speaker's request to assign him/herself (speaker) to speaker and (b) speaker's request to assign side-participant to addressee

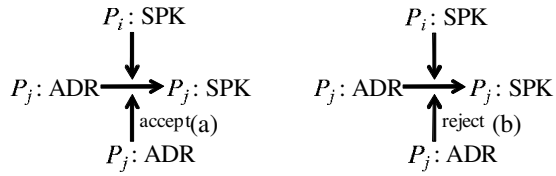


Figure 6-4: Example of answer: (a) addressee's acceptance of speaker's request to assign addressee to speaker and (b) addressee's rejection of speaker's request to assign addressee to speaker

1. `REACT_TO_ALMOST`: function to react to MC's statement "Almost"
2. `REACT_TO_CORRECT`: function to react to MC's statement "Correct"
3. `HESITATE`: function to hesitate to answer (to say something when the answer is not known)
4. `MUTTER`: function to mutter (to hint)

Functions 1 and 2 are reactive behaviors in response to the MC's specific actions. Function 3 lacks the substance of the `ANSWER` and `INFORM` behaviors in the previous section. Function 4 is independent of the game's progression but depends on the question.

6.3 System Implementation

6.3.1 Situation Understanding

At any given moment, the system must select an action that satisfies all the constraints described in the previous section. Ideally, this action selection should be continuous. However, the robot cannot physically perform several actions in a limited time. In order to satisfy the constraints described in section 6.2, the robot must answer when it is asked and should look at another participant when he or she is responding. Thus, the robot must synchronize these external events. In order to activate communication, the system cannot be passive or silent when these external events fail to happen. In this system, we define two kinds of triggers that generate the action selection: an *external trigger* and an *internal trigger*. This system overview is shown in Figure 6-5.

Speech Recognizer, Image Processor & Environmental Information Manager

The *environmental information manager* manages the robot's external information as a situation, and it generates an external trigger. There are many ways to obtain the external information. In this study, we use speech recognition for the MC's speech and image processing for the other panelists' faces.

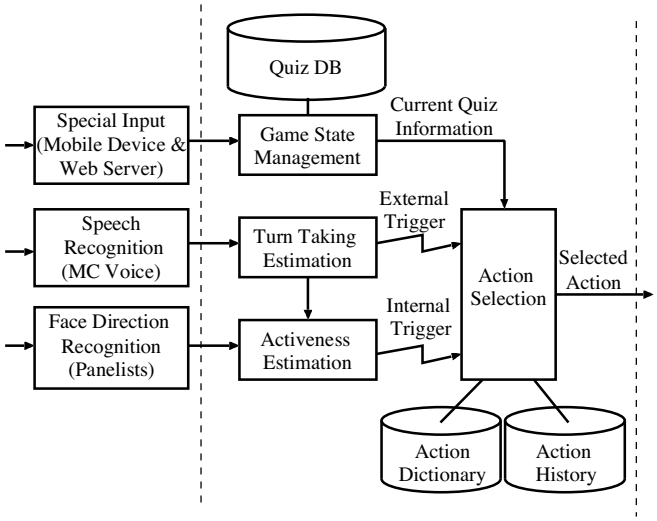


Figure 6-5: System flow of action selection

The MC's speech plays an important role in understanding the situation. For example, when the MC encourages a panelist to answer, that panelist is expected to respond within a short period of time. When the MC says "It's close but not correct" as an evaluation of a player's answer, the system can assume that one of the panelists answered. We believe that the MC's speech is the only relevant speech in a quiz game and that he or she should be able to use a microphone without creating unnecessary stress. Thus, in this system, the MC wears a head-set microphone, enabling speech recognition. We use SKOOD (Shibata and Kobayashi, 2001), a speech recognizer developed by our group.

In image processing, we recognize panelists' facial directions and expressions using images captured by a camera mounted on the robot's eye. In this study, the face extraction is performed using a cascade of boosted classifiers that work with Haar-like features (Viola and Jones, 2001) provided in OpenCV[1]. After the images are extracted, we apply active appearance models (Matthews and Baker, 2004) that are fitted to the extracted region. Using the shape variation parameters of active appearance models, we approximate the direction of each panelist's face and determine whether he or she is looking at the MC, the whiteboard, the robot, or nothing relevant, which means he or she is not actively participating. Moreover, when information for the same face is captured, we recognize estimated changes in facial expressions by calculating the amount of change in a given region. This simple method does not enable us to precisely understand facial expressions, but we assume that any changes in facial expressions are smiles or possess some positive meaning.

External Trigger

External triggers are generated when the MC sets a quiz, encourages one of panelists to answer, or evaluates an answer from a panelist. Upon receiving these triggers, the robot correspondingly reacts to the given quiz, looks at the answering panelist, or answers when appropriate. The MC's speech plays an important role

[1] http://opencv.willowgarage.com/

in understanding the situation. For example, when the MC encourages a panelist to answer, that panelist is expected to answer within a brief period of time. When the MC says "It's close but not correct" as an evaluation of someone's answer, the system assumes that one of the panelists answered. We believe that the MC's speech is the only relevant speech in a quiz game and that he or she should be able to use a microphone without creating unnecessary stress. Therefore, in this system, we equip the MC with a head-set microphone, and the system uses it to perform speech recognition. We use SKOOD (Shibata and Kobayashi, 2001), a speech recognizer developed by our group.

Internal Trigger

Internal triggers are generated when the activeness of a given situation drops below a predefined threshold. Using this trigger, the system takes an action, even though it is not explicitly triggered by an external trigger. In this study, we define activeness as the number of answers given and each panelist's attitude toward participation.

In order to estimate each panelist's activeness, we recognize the panelists' facial directions and expressions using images captured by a camera mounted on the robot's eye. We estimate the direction of each panelist's face and determine where the panelist is looking: at the MC, the whiteboard, the robot, or nothing, which indicates the panelist's lack of active participation. Moreover, when information for the same face is captured, changes in facial expressions are loosely recognized by calculating the amount of change in a region. This simple method does not enable us to precisely understand the facial expressions, but we assume that any changes in facial expressions represent positive emotions.

State Manager

In quiz games, not only are there terms for responses but also terms that allow the MC to give comments about a question and the panelists to give impressions about the question after the correct answer is given. In this system, these terms are separated as states. The set of behaviors from which the robot selects a behavior changes depending on the current state. Although the special states, *initial state* and *final state*, are defined, there are only two essential states: *game state* and *correct state*. The game state is the state in which the panelists should answer, and the correct state is the state that follows a correct response. The behavior sets corresponding to these states are defined in the database. The transition between states occurs in specific situations, such as when the external trigger (the MC saying "Yes, it's correct" in response to someone's answer) is generated or when the MC selects a new quiz using his or her mobile device.

6.3.2 Behavior Evaluation

The behavior evaluation group begins to progress when each state manager has generated triggers. It evaluates all the behaviors in the behavior dictionary and returns the behavior with the highest value, transferring it to the output group. According to the predefined rules, each state manager generates triggers. The multiparty conversation state manager generates its triggers when participation roles change and participants request or answer a question. The game state manager generates its trigger when the game situation changes. The activation state manager generates its trigger when the panelists' communication activeness is lower than the threshold.

The behavior evaluation flow is shown in Figure 6-6. Behavior evaluation begins with the behavior calculating functions. Although most of the behavior functions are statistically predefined, as with the multiparty conversation perspective, the functions are determined based on the situation and using partially predefined

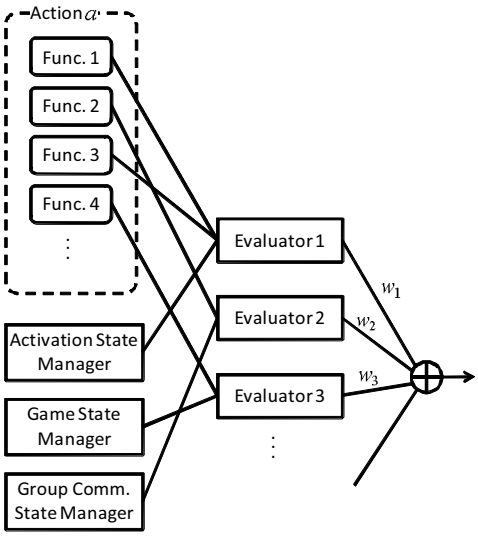


Figure 6-6: Behavior evaluation: after the system calculates functions of each behavior, each evaluator calculates a value based on the optional situations and functions, where the evaluation value of each behavior is the sum of the weighted values.

functions. After this process is complete, the system evaluates the behavior value using several functions and the situation in each perspective. The evaluation process progresses using multiple evaluators. Each evaluator returns values utilizing optional information in the state managers and functions of each behavior. The evaluation returns simple true or false values or a real number. If a true or false value is returned, evaluators perform the evaluation based on rules, such as: "As it concerns the game's progress, the current situation suggests that the MC ask the robot for an answer. Therefore, this behavior has the function `ANSWER`." When real-number values are returned, the evaluation result changes continuously. This case depends on the communication activation. For instance, the "Ask panelist A for an answer" behavior is not very effective when A has a high activeness value. However, when A has a low activeness level, the behavior should be effectively modified, resulting in the following value.

$$e = \begin{cases} \dfrac{a_{\text{MAX}} - a_A}{a_{\text{MAX}}} & \text{if} \quad a_A < a_{\text{MAX}} \\ 0.0 & \text{otherwise} \end{cases}$$

Here, a_A is A's activeness and a_{MAX} is the maximum expectation of the asking behavior's execution, which is predefined. The final evaluation value of each behavior is a weighted sum of the evaluators.

An example of a true or false evaluation is shown in Table 6.1. In this example, the evaluator adheres to two viewpoints: (1) the multiparty conversation perspective—the robot should not assign participants to the role of bystander, and (2) the game progress perspective—the robot should reply to the MC's requests and should not disrupt the game's progress. This is independent of the activation perspective. Moreover, it is only one of the multiple evaluators used in this framework. System designers can easily add evaluators to the system in order to improve the robot's behavior.

Table 6.1: Example of evaluator that calculates using true and false values

Point of view	Weight	Situation	Function
Multiparty conversation	-100.0	$*$	Request to assign other side participant to bystander
Multiparty conversation	$+100.0$	Optional request	acceptance to the request
Nandoku Game	$+100.0$	Pre-answering state MC's request of robot to answer	`ANSWER`
Nandoku game	-100.0	Pre-answering Other panelist is answering	`ANSWER`
Activation	$+100.0$	Pre-answering MC's statement "Almost""	`REACT_TO_ALMOST`
Activation	$+100.0$	Pre-answering MC's statement "Correct""	`REACT_TO_CORRECT`

6.3.3 Sentence Generation

Topic Tracing

In order to intentionally facilitate a conversation, we must define the "topics" in the system. For instance, if the answer to a quiz question is "Hollywood," then "movie" may be a related topic. Similarly, "Roman Holiday" is also related to "movie." Each statement is linked to one of the topics. Topics are transitioned by the robot itself. One topic can transition into multiple topics. For instance, if the current topic is "movie" and the robot says "Speaking of movies, I love Roman Holiday," then the current topic has transitioned to "Roman Holiday." Figure 6-7 shows an example of a topic tree diagram.

The dialogue manager contains all the defined topics, the current quiz question, the current topic, the expected proceeding topics, the current predicate, the current trigger conditions, a history of the topics, a history of the expected topics, a history of the dialogue, a history of the chatting actions, and many other relevant data. The fading coefficient of the previous "next topic" is predefined. We define it as 0.5. For instance, if the current topic is "movie" and the robot says "I love Titanic," then the next topics will be "Titanic" and "movie" with respective scores of 1.0 and 0.5.

Answering Questions

Although the topic information is sufficient to allow the robot to produce its statements, it is not a dialogue. It is a monologue. Therefore, we define relationships between the topic and the utterances as follows.

1. Question Types: 5W1H interrogatives (who, what, how, etc.)

2. Predicate: verbs and adjectives

For instance, if the system estimates the current topic to be "movie," and a user asks "What is your favorite?" the robot can answer "Roman Holiday." Each topic trigger includes three aspects: keywords, the current speaker ID (MC, A, B, or C), and the optional time restriction. Trigger keywords are expressed by a list of disjunctive conditions, including question types and predicates. Examples of this are {"who," "like," "favorite"}, {"what," "which," "like," "favorite"} and {"when," "see"}.

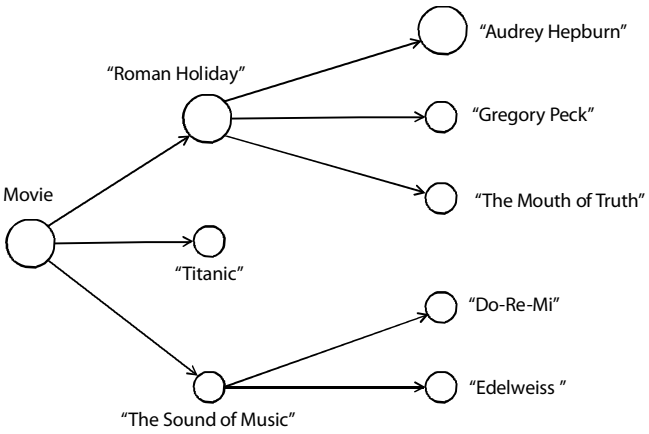


Figure 6-7: Topic tree—tree's size representing number of sentences

Dialogue Actions

Dialogue actions for the system fall into three categories: solving actions, chatting actions, and generic actions. All of them contain a response, an ID, an action list (Robot's body and facial actions), a list of next topics, and a linked action. The linked action provides a mechanism that issues multiple actions in one turn, which produces a combination of responses and spontaneous utterances. This mechanism is described in **6.3.3**.

Solving Action: Solving actions contribute to progressing the game's tasks. These actions can be separated into three types, indicating a "hint," an "answer," or an "initial" chatting action. All three of these exist in the pre-answering state. The data structure of the solving actions is shown in Table6.2. The "initial" chatting action is the first action performed in the post-answering state. For instance, when the MC says "That's correct. It's Hollywood. What do you think about Hollywood, ROBISUKE?" the robot may say "What's more Hollywood than movies?" This expression is the "initial" chatting action.

Chatting Action: Chatting actions expand topics that contain the information defined in **6.3.3**. The data structure of the chatting actions is shown in Table 6.3. It includes the topic, predicate, question type, passive response, active response, and next topic. Passive responses are answers to questions that the other participants might ask. Active responses are spontaneous responses. The mechanism for combining passive and active responses is described in the following section.

Generic Action: Generic actions include responses to praise or statements like "I don't know." A generic action is independent of the topic or question. The generic action data structure is shown in Table 6.4.

Utterance Combination

The response selection process depends on the game state. **Pre-answering state:** If the trigger is a low-activeness trigger, the dialogue manager generates a hint or answer-solving action. The type, hint or answer, is randomly selected. If the trigger is a topic trigger, a chatting action may be generated. The probability of producing a chatting action is $exp(0.5 * chattinglevel)$. The chatting level increases when a consecutive chatting action is chosen, and resets to 0 if another action is chosen. The reason for this is that chatting

Table 6.2: Solving action items

Items	Meaning
Type	”hint” (only for the pre-answering state)
	”answer” (only for the pre-answering state)
	”initial” (only for the pre-answering state)
Response	question-answering utterances
Next topic	expected next topics

Table 6.3: Chatting action items

Items	Meaning
Topic	name of topics
Predicate	verbs or adjectives
Question type	5H1H interrogatives
Passive response	question-answering utterances
Active response	spontaneous utterances
Next topic	expected next topics

Table 6.4: Generic action items

Items	Meaning
Type	”dont_know”
	”impress”
	”react_to_praise”
Response	generic utterances depend on types

is not always necessary in the pre-answering state and the opportunity to respond needs to be given to the other participants. When an action is selected, the dialogue manager notifies the behavior evaluator and increments the hint/answer count by one if the chosen action is a solving action. If the trigger is a low-activeness trigger, the dialogue manager does nothing, and the behavior evaluator selects a generic action with an utterance, such as ”Give us a hint.” These expressions are randomly selected.

Post-answering state: If the trigger is a low-activeness trigger, the dialogue manager generates an active response—a chatting action. An active chatting action is selected by searching all the next topics in descending order based on their scores and randomly choosing an unused chatting action with an active response under the first proceeding topic that contains an unused active chatting action. If the trigger is a topic trigger, a passive response or a combination of a passive and active response is selected. When an action is selected, the dialogue manager notifies the behavior evaluator of the action and increments the hint/answer count by one if the chosen action is a solving action.

There is a probability that each chatting action passive response is linked with an active response, which is determined by multiple factors. In order to adjust the probability of linking passive and active responses, we design a feedback mechanism. The robot is rewarded when participants praise the robot's specific statements. For instance, when the robot says ”My favorite movie is Roman Holiday” and one of participants replies ”That's great” in praise of the statement, the robot should learn that this statement is acceptable in this kind of situation. Active responses appear in two situations: 1) when a low-activeness trigger is received, or 2) when a passive response is generated.

In the first case, the probability of outputting an active response is 1.0, except when all possible active responses have been exhausted. Active responses are randomly selected, and their probability is proportional to the link score recorded in the database. The second case is more complex. First, the dialogue

Table 6.5: Example of conversation between an MC and a robot using the proposed system

	Utterance	Current topic	Expected next topic (score of topic)
Robot	"What's more Hollywood than movies?" (initial utterance)	-	"movie(1.0)"
MC	"**What** is your **favorite** movie?" (type="what", predicate="favorite")	-	-
Robot	"Roman Holiday" (passive response) "Because I love Audrey Hepburn" (active response: combination score = 0.8)	"movie", "Roman_Holiday"	"movie(0.5)", "Roman_Holiday(1.0)", "Audrey_Hepburn(1.0)"
MC	"**What** do you **think** about her?" (type="what", predicate="think")	-	-
Robot	"She is like a fairy." (passive response) (active response: no candidates)	"Audrey_Hepburn"	"fairy(1.0)", "Audrey_Hepburn(0.5)", "Roman_Holiday(0.25)"

manager selects the best passive response. Then it determines all possible active responses given this passive response. The probability of each active response is proportional to:

1. $score_{passive,active}$—the score of a combination of a passive and an active response and

2. the expectation score of the active action's topic.

If for all possible active actions we have $score_{passive,active} = 0$, then $sum(P(active|passive)) = 1$. The number of positive feedbacks $P(active|passive)$ is set to be 0.8 when all other active responses have a score of zero. This means that when $score_{passive,active} = 0.8$ and for all other active actions we have $score_{passive,other_active} = 0$, then $P(active|passive) = 0.8$.

The learning feedback given to the database is designed to work in this way: when the topic trigger receives praise keywords, such as "great" and "nice," the last two actions are examined. If they are a pair of linked passive-active actions, $score(passive, active)$ increases by one. Also, the scores of all other semantically similar passive-active pairs will likewise increase.

Mobile Device & Web Server

In this system, a specific participant, such as the MC, uses a mobile device to select a quiz. The device displays the list of quizzes as web content, and the MC selects one of them with a simple click action. After the MC selects a quiz, information about the selected quiz is transferred to the back-end system through the web server. In this way, the robot understands that either a quiz has been selected or changed by the MC.

As a platform for the developed communication activation system, we introduce the multi-modal conversation robot ROBISUKE (Fujie et al., 2008). ROBISUKE has a camera on his eye. Using images captured by this camera, the system performs image processing. In addition to eyes, it has eyebrows and a mouth on its face, allowing it to generate various expressions. It also has arms, which allow it to point to the quiz on the whiteboard or wave in order to encourage other panelists to answer.

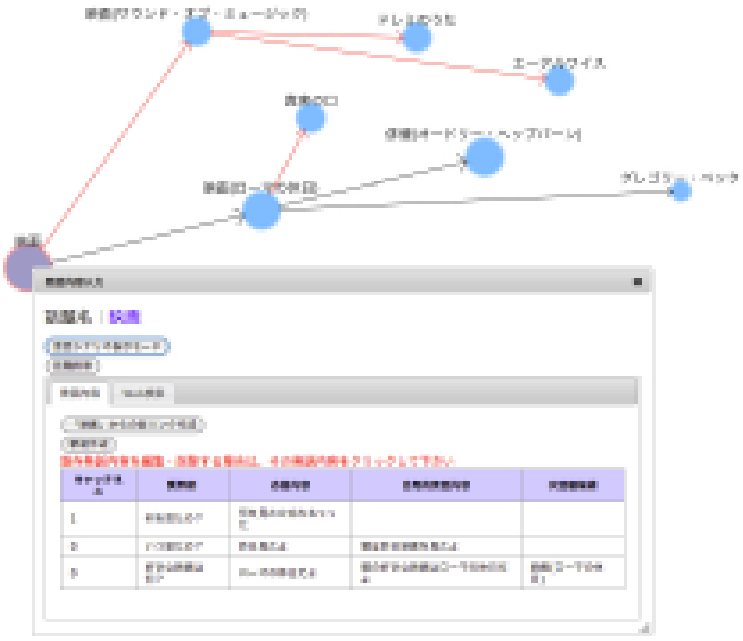


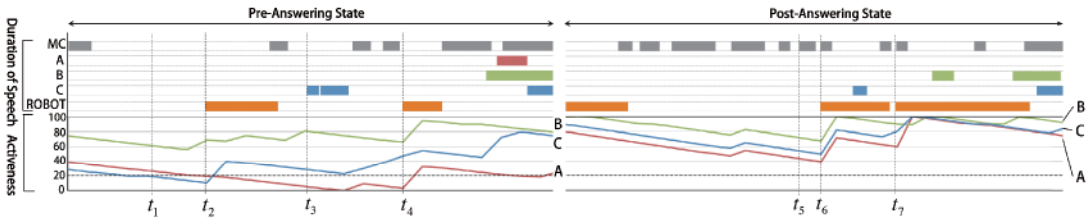
Figure 6-8: Content design tool

6.3.4 Content Design Support Tool

This system contains not only information about the quiz but also answers, monologues, impressions, and stories for the quiz in its database. They are regarded as the contents of the robot, and their extent is one of the important factors used to evaluate the panelists' enjoyment of the robot. While we proposed an automatic sentence generation mechanism in chapter 4, the content used for elderly people should be carefully considered according to their generation and interests. In this chapter, in order to support developers in a robot's content preparation, we implement a registration framework as the web-based application. In this application, developers can easily add behaviors of *answer*, *mutter*, *say impression*, or *say stories*. Because these behaviors require different expressible content for each quiz, they must be explicitly provided in the database. First, developers select a quiz and a behavior from one of these four categories. Then they input the speech content, select a gesture to generate, and synchronize the gesture with the utterance. This simple operation allows developers to easily add robot behaviors. This is one of the features of the proposed system. Figure 6-8 displays the content design support tool. This tool allows developers to implement both a *context-first* design, in which topic sequences are designed first, and a *sentence-first* design, where sentences that are associated with certain topics are filled first. Developers can click on a topic circle in order to expand a window that contains lists of sentences related to the topic.

Execution Example

An example of system execution is shown in **Figure6-10**. An external trigger is generated by the MC's speech recognition when the MC sets a question or someone answers correctly, and the robot will enact a behavior. An external trigger is also generated when the MC explicitly says the robot's name, and the robot



t_1 : Active Trigger (Low Activeness of C) is generated. The ROBOT looks at panelist C.
t_2 : Active Trigger (Low Activeness of A) is generated. At this point, C' s activeness is the lowest. So the ROBOT speaks, "How about you, Mr.C?" (ASK C action)
t_3 : Panclist C speaks, "Hmm, I have no idea."
t_4 : ROBOT replies to a MC' s qustion, "I think it' s MUGITONO (this answer doesn' t make scense)." (ANSWER action)
t_5 : MC asks ROBOT a question, "What' s your favorite menu of McDonald' s?"
t_6 : ROBOT replies to a MC' s question, "I like Fileofish." (CHATTING action)
t_7 : ROBOT combines an utterance as an extra information, "I heard that McDonald' s was founded by McDonald brothers in 1940." (CHATTING action)

Figure 6-9: Excerpt from the experiment (the proposed system)

then selects its behavior, such as answering, asking a question about the question, or asking for a hint. If there is silence for a while, an internal trigger is generated and the robot selects a behavior, such as asking a question about the question, muttering, or asking someone to answer.

6.4 Field Experiment

Our field experiment was designed to test the effectiveness of this system in an elderly day-care center. *Nandoku* quizzes are some of the recreations provided in the care center. This facility is located in a suburb of Tokyo and serves a key role in community care. The subjects were three elderly people as the main panelists, other elderly people as side participants, about ten care staff workers, and four experimenters. In all, about twenty people were in the room. The two experimental conditions were: (1) one of the experimenters played the role of MC, and (2) one of the care staff played the role of MC after some guidance with the system. After the experiment, brief interviews were held. The subjects' facial expressions, surroundings, and the robot's vision were recorded with videos.

A nearly equivalent activation effectiveness was observed between the two conditions. The frequency of the panelists' answers and smiles were observed to be almost the same under both conditions. In particular, the subjects substantially responded to the speech variation for each question. The day-care staff worker, who played the role of MC, easily operated the system with minimal guidance (one minute of instruction) and was engaged in the game. This result shows that the task-based actions and variety of contents, expressions, and gestures contained in the database are important parts of the activation tasks involved in group communication. It also demonstrates that this communication robot system can act as a game medium, which inexperienced players, such as elderly people, can easily use and enjoy.

	(MC selects a question with a mobile device. Question:"牛津")
MC	"The next question is this."
	(External trigger generated)
Robot	(Looking at whiteboard) "What's this!?" (Reacting behavior to question)
	(Elongated silence)
	(Internal trigger generated)
Robot	"Ushitsu?" (Answering behavior)
MC	"Ushitsu, umm... to be simple, yeah, looks like that. But it's not correct."
	(Elongated silence)
MC	"Well, let me see..."
	(Elongated silence)
	(Internal trigger generated)
Robot	"Well, is it related to a cow?" (Asking for a hint behavior)
MC	"Yes, it's related to a cow. I'll give you another hint: it's the name of a place."
	(Elongated silence)
Robot	"Gyu-tsu" (Answering behavior)
MC	"Gyu-tsu? That's not a Japanese word, is it?."
C	"Is the place's name related to a cow?" (Question behavior)
MC	"That's right."
MC	"How about you, SCHEMA?"
	(External trigger generated)
Robot	"Is it a person's name?" (Question behavior
MC	"No, no. It's the name of a place."
	(Elongated silence)
	(Internal Trigger generated)
Robot	"How about you, Mr. B (B's name)?" (Letting B answer behavior)
B	"Umm... I'm not sure."
MC	"Look at the first Kanji character. The first one is not a cow but male..."
	(Elongated silence)
	(Internal trigger generated)
Robot	"I think a cow is for female, an ox is for male..." (Muttering behavior)
C	"I see, that's Oxford."
MC	"Yes. That's correct."
	(External trigger generated)
Robot	"Yeah, I see, that's Oxford!..." (Reacting to correct answer behavior)

Figure 6-10: Example of Nandoku game

Figure 6-11: Scene from the field experiment

6.5 Laboratory Experiment

6.5.1 Experimental Design

Kanda et al. evaluated the effects of a robot's eye gaze on subjects' impressions using an SD method (Kanda et al., 2001). Based on prior methodology, we conducted a video analysis and used the SD method to evaluate how the subjects' impressions changed between two cases: one with robot participation and the other without a robot. We had three subjects (A, B, and C) participating in one game. Thirty subjects participated in the experiment for 20 trials. All subjects were native Japanese speakers and undergraduate or graduate students at Waseda University with a variety if majors. Nine were studying science or engineering; five, political science; five, literature; two, human science; three, education; two, social science; two, law; one, commercial science; and one, liberal arts.

First, subjects were given a brief description of the purpose and procedure of *Nandoku* and told that a robot would participate in the game as a panelist. After the instruction, they were asked to review and sign a consent form. The experiment was conducted with ten groups of three subjects each. Each group played under the following two conditions.

Condition 1: a robot participated in *Nandoku* as a panelist, and an experimenter played the role of MC. The game continued for 15 minutes.

Condition 2: a robot did not participate in *Nandoku*, and an experimenter played the role of MC. The game continued for 15 minutes.

Five of the ten groups played with the robot (**Condition 1**) first. The other five played without the robot (**Condition 2**) first. Two experimenters, with sufficient knowledge of the system, played the role of MC. In order to avoid the effectiveness skewing due to the performance of different MCs, each MC was assigned both conditions in the same group.

Two measurements were calculated in order to evaluate the effectiveness of the communication activation.

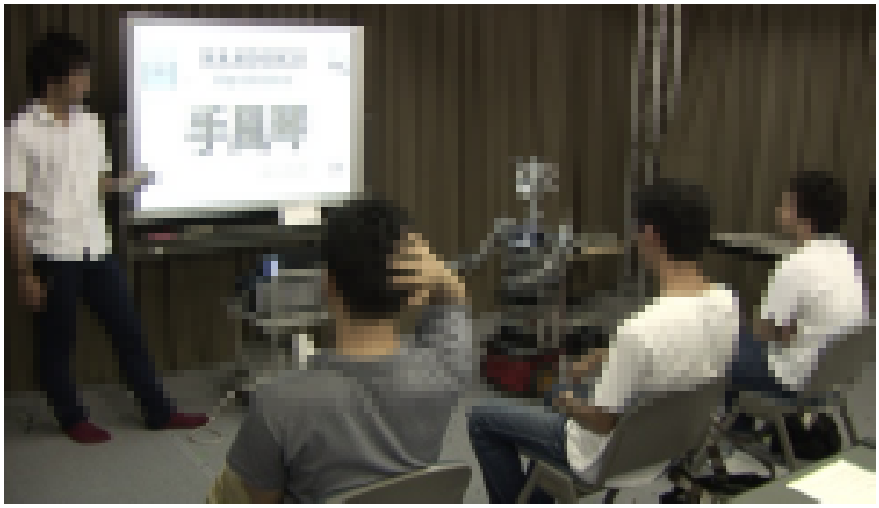


Figure 6-12: Scene of the experiment with robot

Video analysis: We recorded the subjects' behaviors with two video cameras and annotated smiling time for each subject using ANVIL, a video annotation research tool[2]. The annotation of all video was performed by one experimenter in order to avoid the difference in annotation among individuals.

SD analysis: After finishing the game under each condition, subjects were asked to complete a questionnaire that compared their experiences using 30 adjective pairs (in Japanese) with a one-to-seven scale, consisting of "very A," "a little A," "rather A," "not sure," "rather B," "a little B," and "very B," which is based on the SD method. They were also asked to answer another free-form questionnaire about impressions concerning the overall interaction.

For the experimental platform of the developed communication activation system, we introduce the multi-modal conversation robot "SCHEMA" (Matsuyama et al., 2009). SCHEMA is 120 cm tall and has a camera on its eye, which is roughly the same height as the eye of the average seated male. In addition to eyes, it has eyebrows and a mouth, which can simulate various expressions. It also has arms so that it can point to the quiz on the whiteboard and wave in order to encourage other panelists to answer.

6.5.2 Results

The smiling frequency of a subject's face is calculated as $(SmilingTime/TotalTime) \times 100$. The average smiling frequencies in Conditions 1 and 2 are 13.2% and 11.1%, respectively. This result shows that the smiling frequency in Condition 1 was greater than that in Condition 2 for both MCs ($p < .05$).

SD Analysis

Table 6.7 shows the results of the ratings. The adjective pairs in the table are translations of the Japanese words used in the questionnaires. The average values are the results for the two conditions of all thirty subjects. The ratings are based on the one-to-seven scale, where seven implies a strong alignment with the positive adjectives (the adjectives in the leftmost column in **Table 6.7**).

[2]http://www.anvil-software.de/

Table 6.6: Comparison of average and deviation of factors in each condition

	Robot	With Robot			Without Robot		
	MC	A	B	$*$	A	B	$*$
Num. of sub.		15	15	30	15	15	30
Score	pleasure	0.18	0.16	0.17	- 0.55	0.17	- 0.19
	silence	0.49	0.25	0.37	- 0.69	- 0.09	0.39
	ease	- 0.19	0.13	-0.03	- 0.19	- 0.26	0.04
SD	pleasure	1.03	1.13	1.07	1.10	0.78	1.01
	silence	1.29	1.01	1.14	0.83	0.99	0.95
	ease	1.07	1.08	1.07	1.00	1.19	1.10

Note: These are the average and standard deviations of the factor scores. "A" represents the fact that experimenter A played the role of MC, and "$*$" represents the total score and deviation for each condition.

A factor analysis was performed on the SD-method ratings for the 30 adjective pairs. Based on the difference in eigenvalues, we adopted a solution that consists of three factors. The retrieved factor matrix was rotated using the Varimax method in **Table 6.8**. We interpreted the factors by referring to the adjective pairs that have loadings greater than 0.5 in **Table 6.8**. The first factor was named **pleasure factor** because several pleasure adjectives, such as "pleasant" and "cheerful," had a substantial loading on the first factor alone. Therefore, we interpreted this factor to imply that the subjects enjoyed the game. Because calm adjectives, such as "silent" and "orderly," were highly loaded on the second factor, the second factor was named **silence factor**. The third factor was named **ease factor** because these adjectives represent the easiness of the game.

Table 6.6 provides the average and standard deviations of factor scores for each condition.

We considered the significant differences for each condition using a t-test. Our hypotheses are as follows.

Hypothesis 1: Subjects feel more pleased about the game with a robot, regardless of which experimenter plays the role of MC.

Table 6.6 indicates that subjects feel more pleased about the game in which a robot participates when experimenter A is the MC ($p < .01$). There is no significant difference between the two conditions when experimenter B is the MC. Subjects feel more pleased about the game with a robot, although there is no overall significant difference. Therefore, the hypothesis is partially confirmed, although the robot's effectiveness depends on the experimenter's abilities.

Hypothesis 2: Subjects feel the game is noisier when a robot participates, regardless of which experimenter plays the role of MC.

Table 6.6 indicates that subjects feel the game is less silent when a robot participates ($p < .05$), regardless of which experimenter plays the role of MC. In particular, when experimenter A is the MC, there is significant difference ($p < .01$). However, when experimenter B is the MC, there is no significant difference. Therefore, we can conclude subjects feel the game is less silent when the robot participates, meaning they feel the game is noisier with the robot, even though the effect depends on the MC's ability.

Hypothesis 3: Subjects feel the game is easier when a robot participates, regardless of which experimenter plays the role of MC.

Table 6.6 indicates that the subjects' ease factor decreases when the robot participates, although the ease factor maintains a high level when experimenter B is the MC because of experimenter B's expertise in playing the role of MC for this game. Although the ease factor maintains a low level when experimenter B is the MC, the table indicates the ease slightly increases with the robot's participation. Overall, there is no

Table 6.7: Evaluated adjective pairs and results

Adjective Pairs		Avg.(1)	Avg.(2)	S.D.
smooth	rough	4.50	4.87	1.63
natural	unnatural	3.97	5.17	1.54
pleasant	unpleasant	5.53	5.13	1.10
substantial	insubstantial	5.17	5.30	1.11
cheerful	cheerless	5.17	4.97	1.18
warm	cold	4.77	4.77	1.27
friendly	unfriendly	5.27	5.43	1.26
good	bad	5.57	5.37	1.07
open	closed	4.77	4.63	1.15
calm	chaotic	4.73	5.17	1.29
neat	messy	4.23	4.70	1.35
orderly	disorderly	4.20	5.17	1.38
simple	complicated	4.43	4.07	1.36
lively	dull	4.80	4.83	1.42
active	inactive	4.90	4.10	1.31
relaxed	tense	4.70	4.30	1.83
easy	uneasy	4.50	4.63	1.39
carefree	stressful	4.90	5.03	1.45
casual	formal	5.00	4.73	1.44
sociable	unsociable	5.30	4.80	1.13
positive	negative	5.27	5.00	1.02
light	dark	5.37	4.87	0.99
comprehensive	incomprehensive	5.33	5.47	1.20
fun	gloomy	5.93	5.57	1.05
silent	noisy	4.27	5.17	1.22
interesting	boring	5.87	5.43	1.22
attractive	unattractive	5.57	4.93	1.14
light	heavy	5.00	4.57	1.18
conversable	embarrassed	4.80	5.03	1.42
peaceful	anxious	4.80	5.03	1.32

Note: Avg.(1) represents the average for Condition 1, and Avg.(2) represents the average for Condition 2. The average for each condition and the standard deviations represent the ratings of the 30 subjects for the 30 adjective pairs. The ratings are based on the 1-to-7 scale, where 7 strongly adheres to the positive adjectives (shown in the leftmost column).

significant difference between the cases with and without the robot's participation. Therefore, we cannot assert that the robot effectively eases the game with its participation.

From the free-form questionnaires, we can determine that this is due to the repetition of the robot's statements and their inappropriate timing because these issues violate the constraints of group communication.

Free-Form Questionnaire

In the free-form questionnaire, we asked the subjects the following four questions in order to clarify their impressions regarding the robot's participation or absence.

1. Is there any difference between the with-robot and without-robot conditions? Please describe what you were aware of during the interaction.

Table 6.8: Factor matrix (Varimax rotated)

	I	II	III	Communality
pleasant	**0.82**	- 0.05	-0.02	0.67
good	**0.81**	0.11	0.24	0.72
fun	**0.72**	0.10	0.21	0.57
substantial	**0.71**	0.17	0.12	0.56
cheerful	**0.71**	- 0.20	0.12	0.56
friendly	**0.69**	0.32	0.24	0.64
interesting	**0.62**	- 0.07	0.23	0.45
warm	**0.59**	0.11	0.29	0.44
attractive	**0.56**	0.28	- 0.42	0.57
open	**0.53**	- 0.27	0.37	0.49
light	**0.52**	- 0.33	0.39	0.52
silent	- 0.15	**0.71**	- 0.09	0.53
calm	0.14	**0.64**	0.03	0.43
neat	- 0.15	**0.60**	0.11	0.40
orderly	0.10	**0.58**	0.31	0.44
easy	0.24	- 0.03	**0.77**	0.65
casual	0.47	- 0.02	**0.75**	0.79
lively	0.47	0.04	**0.64**	0.63
carefree	0.30	0.03	**0.63**	0.49
active	0.30	- 0.20	**0.60**	0.48
smooth	0.01	- 0.03	**0.60**	0.34
relaxed	- 0.01	0.10	**0.57**	0.33
airy	0.32	- 0.04	**0.55**	0.41
sociable	0.48	- 0.15	**0.51**	0.52
natural	0.14	0.22	**0.51**	0.33
conversable	0.29	0.16	0.50	0.36
comprehensive	0.15	0.21	0.48	0.29
positive	0.45	- 0.27	0.43	0.47
peaceful	0.19	0.41	0.35	0.33
simple	0.02	0.11	- 0.03	0.01
Variance	6.38	2.53	5.51	

Note: The factor matrix is obtained using factor analysis and Varimax rotation. These factors (I to III) were interpreted by referring to factor loadings over 0.5 (shown in boldface) and were respectively named pleasure, silence, and ease.

2. Please describe what you felt, both good and bad, about the robot's behaviors.

3. What other kinds of tasks should the robot perform? Please describe any ideas you may have.

4. Any other comments

Comments supporting Hypothesis 1 (pleasure factor): "The robot's active verbalization made the situation livelier." "The robot's speech with an electric sound felt humorous to me. Also, its statements about trivia and irrelevant answers made us laugh a lot."

Comments supporting Hypothesis 2 (silence factor): "It was completely silent without the robot when the three subjects were silent, but the robot somehow prevented it." "It seemed that the subjects spoke more with the robot present."

Comments supporting Hypothesis 3 (ease factor): "The robot's vocalizations made it easier to speak and express my emotions." "I felt more at ease to laugh and less reserved."

Negative comments: ”I felt the without-robot situation was more natural. The robot made the game feel stricter.” ”I felt it was more difficult to be aware of my own time to speak with the robot.”

These results from the free-form questionnaire and the SD method support each hypothesis. However, the activity inhibitions that resulted from the robot's violation of the group communication constraints still remain. From the comments, we can assume that this is because of the bad timing of the robot's behavior. Therefore, we will analyze the timing of the behavior for each situation as a next step.

6.6 Conclusions and Future Work

We proposed a system that can participate in and activate a quiz game, and we implemented it on a conversation robot. By regarding the robot as a participant in the game, the proposed system enabled the robot to activate the game. This is a new approach to using a robot, which focuses on the fact that a panelist can initiate communication as effectively as a good MC.

We evaluated the communication activation effectiveness using video analysis and an SD-analysis method. As a result of the SD analysis, the subjects were more pleased and felt the game was noisier when the robot participated. This evidences the robot's communication activation function in the party game. However, the ease factor in the game decreased, and that limited the system's effectiveness. From the free-form responses, we believe this problem is caused by a violation of the group communication constraints.

"The medium is the message."

Marshall McLuhan

"I believe in being an innovator."

Walt Disney

7 Conclusions

7.1 Summary of the Dissertation

In this dissertation, we studied computational model of facilitation process in multiparty situations. In **chapter 2**, we proposed the SCHEMA Framework, a computational architecture for multiparty conversation facilitation robots, based on literatures about multiparty conversations and small groups, which have been discussed in domains of social psychology, linguistics and cognitive science. The architecture consists of situation understanding, procedural behavior production and language generation processes. Cognitive process of conversations refers not only declarative memories, but also procedural memories that is independent from semantic representations. Procedural process, the former part in the architecture, includes turn-taking, addressing and engagement control. Language generation process, the latter part, takes into account both verbal and nonverbal language planning. These sub-processes were discussed in details in chapter 3 and 4.

In **chapter 3**, we proposed a framework for conversational robots facilitating a small groups, formalizing with a four-participant group as the smallest unit of facilitation model. We presented a model of procedures obtaining conversational initiatives in incremental steps to maximize total engagement of such four-participant conversations. These situations and procedures were modeled and optimized as a POMDP. This procedural behavior production module is at the core of the SCHEMA Framework. As the results of two experiments, usages of procedures obtaining initiatives showed evidences of acceptability as a participant's behaviors, and feeling of groupness. As for timings, initiating the procedures just after the second or third adjacency pair parts is felt more appropriate than the first pairs by participants.

In **chapter 4**, as the language generation process of the SCHEMA Framework, we presented the SCHEMA QA, an enjoyable question answering pipeline that has capabilities of expressive opinion generation and additional phrasing mechanisms. The opinion sentences are generated from a large number of reviews in the web. After opinion extraction and sentence style conversion process, opinion candidates are ranked in terms of contextual relevance, length of sentences, and frequency of adjectives. Conven-

tional question answering systems mostly have been focused on functional interactions to achieve specific tasks, therefore, they have been based on the Grice's cooperative conversation principles. However, they are not enough to attract users to engage with the system. In this chapter, we assumed informative productions in our daily enjoyable conversations are appeared by an interlocutor's original way of expressions and viewpoints, which can be represented as frequency of adjectives. Beyond simple exchange of questions and answers just like most current both academic and industrial question answering systems, this expressive opinions and additional phrasing mechanisms could indicate possibility to trigger users' motivations to continue to interact with systems over a period of time. We conducted two experiments to evaluate enjoyment of opinion generation and additional phrasing mechanisms. The results showed that both were effective to promote users' enjoyment and interests.

In **chapter 5**, we presented the SCHEMA, a robotic platform to implement the whole proposed architecture. The SCHEMA platform includes both a robotic hardware and a software framework. Based on our consideration of conversational protocols sharing with participants in a group, we design the specifications of the mechanical design, allowing facial expressions, head gestures, and directional control of torso, etc. Its exterior has been carefully designed to realize user-friendly styling for all generations from children to elderly people. We also presented and discussed its network protocols, including lower messaging middleware and higher levels of conversational protocols among modules.

In **chapter 6**, we proposed a system design of a party game robot system for elderly care, which can participate in a quiz game as one of the panelists to entertain others. This is a brand new approach with a robot by focusing on the point that a nifty panelist can promote communications among panelists, as well as help a MC to coordinate a game. We evaluated effectiveness of communication activation using video analysis and SD analysis method. As a result of SD analysis, subjects felt more pleased and more noisy with participation of a robot. That implies evidence of the robot's communication activation function in a party game.

7.2 Significant Contributions

In this dissertation, we believe our contributions are at least three things: (1) combining spoken dialogue systems and human-robot interaction research domains, (2) computational modeling of facilitation strategies, and (3) promising and practical applications of facilitation robots.

Combining Spoken Dialogue Systems and Human-Robot Interaction Research Domains

As we reviewed in the introduction of this dissertation, this work attempted to combine both technological backgrounds of spoken dialogue systems and human-robot interaction research fields. In order to create a physically situated robotic system that can participate in a group conversation, it should obey conversational protocols commonly shared in human conversations. Possessing embodiment functionally equivalent to human is essential to realize a natural conversation. For that purpose, we considered and developed a whole architecture of a conversation robot system, including specifications of a hardware level of design, a middleware level of networking protocol design and a higher level of conversational protocols including procedural decision making to obtain an initiative controlling a group. Such a multidisciplinary endeavor never has been attempted in the histories of the both research domains, and may be influential in other domains, such as cognitive science, social science and other related fields.

Computational Modeling of Facilitation Strategies

We attempted to create a computational model of facilitation processes, not only turn-taking phenomena in multiparty, but also procedural strategies to control conversational situations to regulate whole conversational opportunities. While two-participant conversation models have been traditionally considered, a three-participant situation, minimum unit of multiparty conversation, takes on quite different aspects. Imbalance of engagement density is one of major features of it. However, such a social problem can not be autonomously solved by own from inside the multipary situation because understandings of participant structure are usually diverged by each viewpoint, even if a speaker is regarding that he/she is designing utterances addressed to all participants. A facilitator, the fourth participant is the first person who can observed such situations and detect social problems. Therefore, we considered a four-participant model as a facilitation process model. In order to control regulate a socially imbalance situation, we employed a notion of engagement density as a measurement of amount of communication among participants, and considered procedural steps to obtaining an initiative for floor control and topic shift. Also, we considered strategies of language generation style control in terms of an amount of information and length of an utterance.

Promising and Practical Applications of Facilitation Robots

As a promising and practical application of facilitation robot, we proposed a party game system for elderly care. We conducted a real field experiment in which a prototype system participates in an elderly day-care center. The robot's actions and utterances did trigger the elders' and care staffs' big laughs. The conversation continued about an hour until we stopped the system because all the utterance data we prepared was used up. That field experiment implies that such a facilitation robot has a big potential for an entertainment medium, not just because it could entertain the elders themselves, but also it could promote enjoyable conversations between the elders and care staffs.

7.3 Future Work

Modeling a group facilitation process is challenging in both computational and neuroscientifical manners and contexts. Furthermore, combining these research fields must accelerate to deeply understand human's social activities and it would help to build new kinds of human interfaces. Also, a facilitation robot could have a potential of a new type of robotic entertainment medium. There are at least the following attractive research topics.

Rapport in a Group

While this work dealt with engagement density control and language generation, it still lacks deeper mental states in group process. In terms of a feeling of connection and closeness with others, rapport has been identified as an important function of social interaction. It has been reported that rapport has powerful effects on performance in a variety of domains, including negotiation (Drolet and Morris, 2000), counseling (Kang et al., 2012) and education (Bernieri and Rosenthal, 1991). Spencer-Oatey defined, "Rapport refers to the relative harmony and smoothness of relations between people, and rapport management refers to the management (or mismanagement) of relations between people." And she categorized major factors of rapport are *face management*, *mutual attentiveness* and *coordination* (Spencer-Oatey, 2005). Cassell et al. presented their computational dyadic model of rapport based on the Spencer-Oatey's categolization (Zhao

et al., 2014), and proposed a computational architecture for an embodied conversational agent (Papangelis et al., 2014).

As Bernieri wrote, "rapport is a social construct that must be defined at the level of a dyad or larger group" (Bernieri and Gillis, 2001), a dyadic rapport model could be extended to a group rapport model. As we found in Chapter 3, each group has its own characteristic. There are some socio-psychological analyses of stereotypes of groups and their group norms (Adams and Marshall, 1996; Terry et al., 1999; Hogg and Reid, 2006; Christensen et al., 2004). However, there is never a computational model of rapport of a group. While some parts of the dyadic rapport model could be applied to even a group model, group processes should be different from dyadic interactions. We will consider computational models and architectures of rapport in a group.

Neural Basis Cognitive Modeling of Group Interactions

Gazzaniga introduced the term "social brain" into neuropsychology in his studies of emotional and social communication disturbances after righthemisphere damages (Gazzaniga, 1985). Since then the social brain have meant how the human brain processes social information and regulates the mind as a whole. Brothers pointed out that there was a circumscribed set of brain regions that were dedicated to the social brain: amygdala, orbital frontal cortex and temporal cortex (Brothers et al., 2002). Dunbar, an anthropologist, proposed the social brain hypothesis (Dunbar, 1998), where he argued that human intelligence did not evolve primarily to solve ecological problems, but it evolved to survive and reproduce in large and complex social groups. In fact, some of the behaviors have been pointed out to be associated with group oriented behaviors, such as reciprocal altruism, deception and coalition formation. These group oriented behaviors relate to the "Theory of Mind" (Premack and Woodruff, 1978) and the simulation theory, each of which relies on the activity of mirror neurons (Di Pellegrino et al., 1992).

Cognitive developmental robotics is aiming to understand how human's higher cognitive functions by means of a synthetic approach using physical embodiment structuring information through interactions with the environment (Asada et al., 2009). But it has not yet dealt with conversational level, but interaction level. There are also some early stages of research augmenting an agent's ability of multiparty interaction using neural basis information. For example, Ehrlic et al. proposed a social engagement recognition method using brain activities via electroencephalography (EEG). They reported such information is helpful to understand the engagement status (Ehrlich et al.).

While the fields of social neuroscience and cognitive robotics are arising, and there are already sophisticated machine learning methods, few research on computational modeling of conversational agents fueled by deeper understanding of neural basis phenomena has been attempted. We believe such interdisciplinary research of group conversational interactions would help to deal with the complexity in social contexts.

Robotic Entertainment Media

As we presented in chapter 6, conversational robots that have capabilities to facilitate multiparty situations can be robotic entertainment media. Our facilitation robot plays a role of the forth player of our society, which controls its own behaviors to contribute to decide the boundary of ratification of a conversational role. Sometimes the robot would be a bystander only observing a situation, and sometimes it would aggressively obtain an initiative of a conversation. It is a truly new medium worth to think its sophistication in our society. Marshall McLuhan proposes that a medium could affect the society where it plays a role not only by the content delivered over the medium, but also by the characteristics of the medium itself (McLuhan, 1994). A conversational robot should also involve its own message in it as a medium with its own characteristics of

communication protocols. As we discussed in this dissertation, it is reasonable that a conversational robot has its embodiment functionally equivalent to human ways to realize a natural communication. And the way of content deliveries should be realized along conversational protocols shared with human interlocutors.

Content is king for especially entertainment media. AIBO (Fujita, 2000, 2001), QRIO (Ishida, 2004; Ishida et al., 2001) and Paro (Wada and Shibata, 2006) were early entertainment robotic products while they did not have sophisticated conversational capabilities, but lifelike reactive movements. Conversational robot media requires taking both linguistic and non-linguistic contents into account. And as we discussed in chapter 4, the novelty of utterances cannot be defined without interlocutors' user models. Content design for such media is more complex than other existing media.

How can we create more attractive conversations? Can we build an eternally augmented enjoyable conversational system? Such questions cannot be solved only by a single scientific method. It needs combinations of arts and sciences. Creating a robotic entertainment system would be a new media art in this century.

A

POMDP Model Specification

Participants' Actions

	A1OCTF	A2OBFF	A3OBTF	AOOAFF
none	A1OCFT	A2OCTT	A3OBFT	AOOBTT
A1RATT	A1OCFF	A2OCTF	A3OBFF	AOOBTF
A1RATF	A1ON	A2OCFT	A3OCTT	AOOBFT
A1RAFT	A2RATT	A2OCFF	A3OCTF	AOOBFF
A1RAFF	A2RATF	A2ON	A3OCFT	AOOCTT
A1RBTT	A2RAFT	A3RATT	A3OCFF	AOOCTF
A1RBTF	A2RAFF	A3RATF	A3ON	AOOCFT
A1RBFT	A2RBTT	A3RAFT	AORATT	AOOCFF
A1RBFF	A2RBTF	A3RAFF	AORATF	AOON
A1RCTT	A2RBFT	A3RBTT	AORAFT	B1RATT
A1RCTF	A2RBFF	A3RBTF	AORAFF	B1RATF
A1RCFT	A2RCTT	A3RBFT	AORBTT	B1RAFT
A1RCFF	A2RCTF	A3RBFF	AORBTF	B1RAFF
A1RN	A2RCFT	A3RCTT	AORBFT	B1RBTT
A1OATT	A2RCFF	A3RCTF	AORBFF	B1RBTF
A1OATF	A2RN	A3RCFT	AORCTT	B1RBFT
A1OAFT	A2OATT	A3RCFF	AORCTF	B1RBFF
A1OAFF	A2OATF	A3RN	AORCFT	B1RCTT
A1OBTT	A2OAFT	A3OATT	AORCFF	B1RCTF
A1OBTF	A2OAFF	A3OATF	AORN	B1RCFT
A1OBFT	A2OBTT	A3OAFT	AOOATT	B1RCFF
A1OBFF	A2OBTF	A3OAFF	AOOATF	B1RN
A1OCTT	A2OBFT	A3OBTT	AOOAFT	B1OATT

B1OATF	B3RATF	BOOATF	C2RATF	C3OATF
B1OAFT	B3RAFT	BOOAFT	C2RAFT	C3OAFT
B1OAFF	B3RAFF	BOOAFF	C2RAFF	C3OAFF
B1OBTT	B3RBTT	BOOBTT	C2RBTT	C3OBTT
B1OBTF	B3RBTF	BOOBTF	C2RBTF	C3OBTF
B1OBFT	B3RBFT	BOOBFT	C2RBFT	C3OBFT
B1OBFF	B3RBFF	BOOBFF	C2RBFF	C3OBFF
B1OCTT	B3RCTT	BOOCTT	C2RCTT	C3OCTT
B1OCTF	B3RCTF	BOOCTF	C2RCTF	C3OCTF
B1OCFT	B3RCFT	BOOCFT	C2RCFT	C3OCFT
B1OCFF	B3RCFF	BOOCFF	C2RCFF	C3OCFF
B1ON	B3RN	BOON	C2RN	C3ON
B2RATT	B3OATT	C1RATT	C2OATT	CORATT
B2RATF	B3OATF	C1RATF	C2OATF	CORATF
B2RAFT	B3OAFT	C1RAFT	C2OAFT	CORAFT
B2RAFF	B3OAFF	C1RAFF	C2OAFF	CORAFF
B2RBTT	B3OBTT	C1RBTT	C2OBTT	CORBTT
B2RBTF	B3OBTF	C1RBTF	C2OBTF	CORBTF
B2RBFT	B3OBFT	C1RBFT	C2OBFT	CORBFT
B2RBFF	B3OBFF	C1RBFF	C2OBFF	CORBFF
B2RCTT	B3OCTT	C1RCTT	C2OCTT	CORCTT
B2RCTF	B3OCTF	C1RCTF	C2OCTF	CORCTF
B2RCFT	B3OCFT	C1RCFT	C2OCFT	CORCFT
B2RCFF	B3OCFF	C1RCFF	C2OCFF	CORCFF
B2RN	B3ON	C1RN	C2ON	CORN
B2OATT	BORATT	C1OATT	C3RATT	COOATT
B2OATF	BORATF	C1OATF	C3RATF	COOATF
B2OAFT	BORAFT	C1OAFT	C3RAFT	COOAFT
B2OAFF	BORAFF	C1OAFF	C3RAFF	COOAFF
B2OBTT	BORBTT	C1OBTT	C3RBTT	COOBTT
B2OBTF	BORBTF	C1OBTF	C3RBTF	COOBTF
B2OBFT	BORBFT	C1OBFT	C3RBFT	COOBFT
B2OBFF	BORBFF	C1OBFF	C3RBFF	COOBFF
B2OCTT	BORCTT	C1OCTT	C3RCTT	COOCTT
B2OCTF	BORCTF	C1OCTF	C3RCTF	COOCTF
B2OCFT	BORCFT	C1OCFT	C3RCFT	COOCFT
B2OCFF	BORCFF	C1OCFF	C3RCFF	COOCFF
B2ON	BORN	C1ON	C3RN	COON
B3RATT	BOOATT	C2RATT	C3OATT	

initial State

```
# Harmony State :  Motivation State :  Participants' Action
F : none :  none
```

Discount

```
discount:  0.7
```

System Actions

```
null              qCur              tri               rea
a                 qNew              nod
```

Participants' Action Model

```
# useraction :  dialoguestate :  systemaction :  usergoal' :  usergoal2
:  useraction'
AP : ask :  .* :  (nod|null) :  .* :  .* :  ask 0.2
AP : ask :  .* :  (nod|null) :  .* :  .* :  .1R.* 0.6/39
AP : ask :  .* :  (nod|null) :  .* :  .* :  ..R.* 0.17859/116
AP : ask :  .* :  (nod|null) :  .* :  .* :  .* 0.02/158
AP : ask :  .* :  .* :  .* :  .* :  .* 1.0/314

AP : .1.* :  .* :  (nod|null) :  .* :  .* :  .1.* 0.13/78
AP : .1.* :  .* :  (nod|null) :  .* :  .* :  .2.* 0.82/78
AP : .1.* :  .* :  (nod|null) :  .* :  .* :  .3.* 0.05/78
AP : .1.* :  .* :  rea :  .* :  .* :  .* 1.0/314

AP : .2.* :  .* :  (nod|null) :  .* :  .* :  .1.* 0.07/78
AP : .2.* :  .* :  (nod|null) :  .* :  .* :  .2.* 0.15/78
AP : .2.* :  .* :  (nod|null) :  .* :  .* :  .3.* 0.78/78
AP : .2.* :  .* :  rea :  .* :  .* :  .* 1.0/314

AP : .3.* :  .* :  (nod|null) :  .* :  .* :  .1.* 0.360/78
AP : .3.* :  .* :  (nod|null) :  .* :  .* :  .2.* 0.230/78
AP : .3.* :  .* :  (nod|null) :  .* :  .* :  .3.* 0.41/78
AP : .3.* :  .* :  rea :  .* :  .* :  .* 1.0/314

AP : .O.* :  .* :  (nod|null) :  .* :  .* :  .1.* 0.35/78
AP : .O.* :  .* :  (nod|null) :  .* :  .* :  .2.* 0.55/78
AP : .O.* :  .* :  (nod|null) :  .* :  .* :  .3.* 0.1/78
AP : .O.* :  .* :  rea :  .* :  .* :  .* 1.0/314

AP : .* :  .* :  (tri|a) :  .* :  .* :  .1R.* 0.07/39
AP : .* :  .* :  (tri|a) :  .* :  .* :  .2R.* 0.15/39
AP : .* :  .* :  (tri|a) :  .* :  .* :  .3R.* 0.78/39
AP : .* :  .* :  (qCur|qNew) :  .* :  .* :  .1R.* 0.13/39
AP : .* :  .* :  (qCur|qNew) :  .* :  .* :  .2R.* 0.82/39
AP : .* :  .* :  (qCur|qNew) :  .* :  .* :  .3R.* 0.05/39
AP : none :  .* :  (null|nod|rea) :  .* :  .* :  .* 1.0/314
```

Reward model

```
# usergoal :  usergoal2 :  useraction :  dialoguestate' :  systemaction'
```

```
R: .* :  .* :  .1RN.* :  .* :  a 5
R: .* :  .* :  .1RN.* :  .* :  qCur (-10)
R: .* :  .* :  .1RN.* :  .* :  qNew (-10)
R: .* :  .* :  .1RN.* :  .* :  tri 0
R: .* :  .* :  .1RN.* :  .* :  nod 0
R: .* :  .* :  .1RN.* :  .* :  rea (-10)
R: .* :  .* :  .1RN.* :  .* :  null 0
R: .* :  .* :  .(2|3)RN.* :  .* :  a (-3)
R: .* :  .* :  ..ON : .* :  nod 5
R: .* :  .* :  ..ON : .* :  null 0
R: .* :  .* :  ..ON.* :  .* :  qCur (-10)
R: .* :  .* :  ..ON.* :  .* :  qNew (-10)
R: .* :  .* :  ask :  .* :  rea 10
R: F : .* :  .1R.*[^N] : .* :  a 5
R: F : .* :  .1R.*[^N] : .* :  qCur (-10)
R: F : .* :  .1R.*[^N] : .* :  qNew (-10)
R: F : .* :  .1R.*[^N] : .* :  tri 0
R: F : .* :  .1R.*[^N] : .* :  nod 0
R: F : .* :  .1R.*[^N] : .* :  rea 0
R: F : .* :  .1R.*[^N] : .* :  null 0
R: F : .* :  .(2|3)R.*[^N] : .* :  a (-10)
R: F : .* :  .(2|3)R.*[^N] : .* :  qCur 4
R: F : .* :  .(2|3)R.*[^N] : .* :  qNew 0
R: F : .* :  .(2|3)R.*[^N] : .* :  tri (-10)
R: F : .* :  .(2|3)R.*[^N] : .* :  nod 10
R: F : .* :  .(2|3)R.*[^N] : .* :  rea 0
R: F : .* :  .(2|3)R.*[^N] : .* :  null 0
R: F : .* :  .1O.*[^N] : .* :  a 0
R: F : .* :  .1O.*[^N] : .* :  qCur 0
R: F : .* :  .1O.*[^N] : .* :  qNew 0
R: F : .* :  .1O.*[^N] : .* :  tri 4
R: F : .* :  .1O.*[^N] : .* :  nod 3
R: F : .* :  .1O.*[^N] : .* :  rea 0
R: F : .* :  .1O.*[^N] : .* :  null 3
R: F : .* :  .2O.*[^N] : .* :  a 0
R: F : .* :  .2O.*[^N] : .* :  qCur 0
R: F : .* :  .2O.*[^N] : .* :  qNew 0
R: F : .* :  .2O.*[^N] : .* :  tri 3
R: F : .* :  .2O.*[^N] : .* :  nod 3
R: F : .* :  .2O.*[^N] : .* :  rea 0
R: F : .* :  .2O.*[^N] : .* :  null 3

R: Pre :  .* :  .1R.*[^N] : .* :  a 5
R: Pre :  .* :  .1R.*[^N] : .* :  qCur 0
R: Pre :  .* :  .1R.*[^N] : .* :  qNew 0
```

```
R: Pre :  .* :  .1R.*[^N] : .* :  tri 0
R: Pre :  .* :  .1R.*[^N] : .* :  nod 0
R: Pre :  .* :  .1R.*[^N] : .* :  rea 0
R: Pre :  .* :  .1R.*[^N] : .* :  null 0
R: Pre :  .* :  .(2|3)R.*[^N] : .* :  a 0
R: Pre :  .* :  .(2|3)R.TT : .* :  qCur 10
R: Pre :  .* :  .(2|3)R.TF : .* :  qCur 10
R: Pre :  .* :  .(2|3)R.FT : .* :  qCur 10
R: Pre :  .* :  .(2|3)R.FF : .* :  qCur 10
R: Pre :  .* :  .(2|3)R.TT : .* :  qNew 0
R: Pre :  .* :  .(2|3)R.TF : .* :  qNew 0
R: Pre :  .* :  .(2|3)R.FT : .* :  qNew 0
R: Pre :  .* :  .(2|3)R.FF : .* :  qNew 0
R: Pre :  .* :  .(2|3)R.*[^N] : .* :  tri 0
R: Pre :  .* :  .(2|3)R.*[^N] : .* :  nod 0
R: Pre :  .* :  .(2|3)R.*[^N] : .* :  rea 0
R: Pre :  .* :  .(2|3)R.*[^N] : .* :  null 0

R: Pre :  .* :  .3O.*[^N] : .* :  a 0
R: Pre :  .* :  .3O.*[^N] : .* :  qCur (-10)
R: Pre :  .* :  .3O.*[^N] : .* :  qNew (-10)
R: Pre :  .* :  .(1|2)O.*[^N] : .* :  tri 3
R: Pre :  .* :  .3O.*[^N] : .* :  nod 0
R: Pre :  .* :  .3O.*[^N] : .* :  rea 0
R: Pre :  .* :  .3O.*[^N] : .* :  null 0

R: T : .* :  .1R.*[^N] : .* :  a 5
R: T : .* :  .1R.*[^N] : .* :  qCur (-10)
R: T : .* :  .1R.*[^N] : .* :  qNew (-10)
R: T : .* :  .1R.*[^N] : .* :  tri 0
R: T : .* :  .1R.*[^N] : .* :  nod 0
R: T : .* :  .1R.*[^N] : .* :  rea 0
R: T : .* :  .1R.*[^N] : .* :  null 0

R: T : .* :  .(2|3)R.*[^N] : .* :  a 0
R: T : .* :  .(2|3)R.TT : .* :  qCur 10
R: T : .* :  .(2|3)R.FT : .* :  qCur 10
R: T : .* :  .(2|3)R.TF : .* :  qCur 10
R: T : .* :  .(2|3)R.FF : .* :  qCur 10
R: T : .* :  .(2|3)R.TT : .* :  qNew (-10)
R: T : .* :  .(2|3)R.FT : .* :  qNew (-10)
R: T : .* :  .(2|3)R.TF : .* :  qNew (-10)
R: T : .* :  .(2|3)R.FF : .* :  qNew (-10)
R: T : .* :  .(2|3)R.*[^N] : .* :  tri 0
R: T : .* :  .(2|3)R.*[^N] : .* :  nod 0
```

```
R: T : .* :  .(2|3)R.*[^N] : .* :  rea 0
R: T : .* :  .(2|3)R.*[^N] : .* :  null 0

R: T : .* :  .3O.*[^N] : .* :  a 0
R: T : .* :  .3O.*[^N] : .* :  qCur (-10)
R: T : .* :  .3O.*[^N] : .* :  qNew (-10)
R: T : .* :  .(1|2)O.*[^N] : .* :  tri 0
R: T : .* :  .3O.*[^N] : .* :  nod 4
R: T : .* :  .3O.*[^N] : .* :  rea 0
R: T : .* :  .3O.*[^N] : .* :  null 0

R: .* :  .* :  .* :  .* :  null 10
```

Bibliography

Gerald R Adams and Sheila K Marshall. A developmental social psychology of identity: Understanding the person-in-context. *Journal of adolescence*, 19(5):429–442, 1996.

Christopher Alexander, Sara Ishikawa, and Murray Silverstein. A pattern language: Towns, buildings, construction (cess center for environmental). 1977.

John R Anderson. *How can the human mind occur in the physical universe?* Oxford University Press, 2007.

John R Anderson, Daniel Bothell, Michael D Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. An integrated theory of the mind. *Psychological review*, 111(4):1036, 2004.

Minoru Asada, Koh Hosoda, Yasuo Kuniyoshi, Hiroshi Ishiguro, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida. Cognitive developmental robotics: A survey. *Autonomous Mental Development, IEEE Transactions on*, 1(1):12–34, 2009.

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In *The semantic web*, pages 722–735. Springer, 2007.

Robert F Bales. Interaction process analysis. *Cambridge, Mass*, 1950.

Kenneth D Benne and Paul Sheats. Functional roles of group members. *Journal of social issues*, 4(2): 41–49, 1948.

Frank J Bernieri and John S Gillis. Judging rapport: Employing brunswik's lens model to study interpersonal sensitivity. *Interpersonal sensitivity: Theory and measurement*, pages 67–88, 2001.

Frank J Bernieri and Robert Rosenthal. Interpersonal coordination: Behavior matching and interactional synchrony. *Fundamentals of nonverbal behavior*, page 401, 1991.

Timothy Bickmore and Julie Cassell. Small talk and conversational storytelling in embodied conversational interface agents. In *AAAI fall symposium on narrative intelligence*, pages 87–92, 1999.

Timothy Wallace Bickmore. *Relational agents: Effecting change through human-computer relationships*. PhD thesis, Massachusetts Institute of Technology, 2003.

Bruce J Biddle. Recent development in role theory. *Annual review of sociology*, pages 67–92, 1986.

Bruce Jesse Biddle. *Role theory: Expectations, identities, and behaviors*. Academic Press New York, 1979.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

Dan Bohus and Eric Horvitz. Open-world dialog: Challenges, directions, and prototype. In *Proceedings of IJCAI'2009 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*. Citeseer, 2009a.

Dan Bohus and Eric Horvitz. Models for multiparty engagement in open-world dialog. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 225–234. Association for Computational Linguistics, 2009b.

Dan Bohus and Eric Horvitz. Dialog in the open world: platform and applications. In *Proceedings of the 2009 international conference on Multimodal interfaces*, pages 31–38. ACM, 2009c.

Dan Bohus and Eric Horvitz. Learning to predict engagement with a spoken dialog system in open-world settings. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 244–252. Association for Computational Linguistics, 2009d.

Dan Bohus and Eric Horvitz. On the challenges and opportunities of physically situated dialog. In *AAAI Symposium on Dialog with Robots*, 2010a.

Dan Bohus and Eric Horvitz. Facilitating multiparty dialog with gaze, gesture, and speech. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, page 5. ACM, 2010b.

Dan Bohus and Eric Horvitz. Decisions about turns in multiparty conversation: from perception to action. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 153–160. ACM, 2011a.

Dan Bohus and Eric Horvitz. Multiparty turn taking in situated dialog: Study, lessons, and directions. In *Proceedings of the SIGDIAL 2011 Conference*, pages 98–109. Association for Computational Linguistics, 2011b.

Dan Bohus, Eric Horvitz, Takayuki Kanda, Bilge Mutlu, and Antoine Raux. Introduction to the special issue on dialog with robots. *AI Magazine*, 32(4):15–16, 2011.

Robert P Bostrom, Robert Anson, and Vikki K Clawson. Group facilitation and group support systems. *Group support systems: New perspectives*, pages 146–168, 1993.

Cynthia L Breazeal. *Designing Sociable Robots*. MIT press, 2004.

Eric Breck, Yejin Choi, and Claire Cardie. Identifying expressions of opinion in context. In *Proceedings of the 20th international joint conference on Artifical intelligence*, pages 2683–2688. Morgan Kaufmann Publishers Inc., 2007.

Leslie Brothers et al. The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Foundations in social neuroscience*, pages 367–385, 2002.

Japanese Government Cabinet Office. *Annual Report on the Aging Society*. Cabinet Office, Japanese Government, 2014.

Angelo Cafaro, Hannes Högni Vilhjálmsson, Timothy Bickmore, Dirk Heylen, and Catherine Pelachaud. Representing communicative functions in saiba with a unified function markup language. In *Intelligent Virtual Agents*, pages 81–94. Springer, 2014.

Nick Campbell and Stefan Scherer. Comparing measures of synchrony and alignment in dialogue speech timing with respect to turn-taking activity. In *INTERSPEECH*, pages 2546–2549, 2010.

Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, et al. The ami meeting corpus: A pre-announcement. In *Machine learning for multimodal interaction*, pages 28–39. Springer, 2006.

Justine Cassell. *Embodied conversational agents*. The MIT Press, 2000.

Justine Cassell and Timothy Bickmore. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and User-Adapted Interaction*, 13(1-2):89–132, 2003.

Justine Cassell, Timothy Bickmore, Mark Billinghurst, Lee Campbell, Kenny Chang, Hannes Vilhjálmsson, and Hao Yan. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 520–527. ACM, 1999.

Justine Cassell, Matthew Stone, and Hao Yan. Coordination and context-dependence in the generation of embodied conversation. In *Proceedings of the first international conference on Natural language generation-Volume 14*, pages 171–178. Association for Computational Linguistics, 2000.

Crystal Chao and Andrea L Thomaz. Controlling social dynamics with a parametrized model of floor regulation. 2012a.

Crystal Chao and Andrea Lockerd Thomaz. Timing in multimodal turn-taking interactions: Control and analysis using timed petri nets. *Journal of Human-Robot Interaction*, 1(1), 2012b.

P Niels Christensen, Hank Rothgerber, Wendy Wood, and David C Matz. Social norms and identity relevance: A motivational approach to normative behavior. *Personality and Social Psychology Bulletin*, 30 (10):1295–1309, 2004.

Herbert H Clark. *Using language*, volume 4. Cambridge University Press Cambridge, 1996.

Herbert H Clark and Thomas B Carlson. Hearers and speech acts. *Language*, pages 332–373, 1982.

Herbert H Clark and Edward F Schaefer. Contributing to discourse. *Cognitive science*, 13(2):259–294, 1989.

Giuseppe Di Pellegrino, Luciano Fadiga, Leonardo Fogassi, Vittorio Gallese, and Giacomo Rizzolatti. Understanding motor events: a neurophysiological study. *Experimental brain research*, 91(1):176–180, 1992.

Kohji Dohsaka, Ryota Asai, Ryuichiro Higashinaka, Yasuhiro Minami, and Eisaku Maeda. Effects of conversational agents on human communication in thought-evoking multi-party dialogues. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 217–224. Association for Computational Linguistics, 2009.

Wen Dong and Alex Pentland. Modeling influence between experts. In *Artifical Intelligence for Human Computing*, pages 170–189. Springer, 2007.

Wen Dong, Bruno Lepri, Alessandro Cappelletti, Alex Sandy Pentland, Fabio Pianesi, and Massimo Zancanaro. Using the influence model to recognize functional roles in meetings. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 271–278. ACM, 2007.

Wen Dong, Bruno Lepri, Fabio Pianesi, and Alex Pentland. Modeling functional roles dynamics in small group interactions. 2013.

Aimee L Drolet and Michael W Morris. Rapport in conflict resolution: Accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. *Journal of Experimental Social Psychology*, 36(1):26–50, 2000.

Peter F Drucker. The discipline of innovation. *Harvard business review*, 63(3):67–72, 1984.

Shelli Dubs and Stephen C Hayne. Distributed facilitation: a concept whose time has come? In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 314–321. ACM, 1992.

Robin IM Dunbar. The social brain hypothesis. *brain*, 9(10):178–190, 1998.

Stefan Ehrlich, Agnieszka Wykowska, Karinne Ramirez-Amaro, and Gordon Cheng. When to engage in interaction-and how? eeg-based enhancement of robot ' s ability to sense social signals in hri.

Nobutsuna Endo, Shimpei Momoki, Massimiliano Zecca, Minoru Saito, Yu Mizoguchi, Kazuko Itoh, and Atsuo Takanishi. Development of whole-body emotion expression humanoid robot. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 2140–2145. IEEE, 2008.

Giuseppe Fabbrizio, Ahmet Aker, and Robert Gaizauskas. Summarizing online reviews using aspect rating distributions and language modeling. *IEEE Intelligent Systems*, 28(3):28–37, May 2013a. ISSN 1541-1672. doi: 10.1109/MIS.2013.36. URL `http://dx.doi.org/10.1109/MIS.2013.36`.

Giuseppe Fabbrizio, Ahmet Aker, and Robert Gaizauskas. Summarizing online reviews using aspect rating distributions and language modeling. *IEEE Intelligent Systems*, 28(3):28–37, 2013b. ISSN 1541-1672. doi: http://doi.ieeecomputersociety.org/10.1109/MIS.2013.36.

Giuseppe Fabbrizio, Amanda J Stent, and Robert Gaizauskas. A hybrid approach to multi-document summarization of opinions in reviews. In *8th International Natural Language Generation Conference - INLG2014*. Association for Computational Linguistics, 2014.

Christiane Fellbaum. Wordnet: an electronic lexical database. 1998. *WordNet is available from http://www. cogsci. princeton. edu/wn*, 2010.

David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A Kalyanpur, Adam Lally, J William Murdock, Eric Nyberg, John Prager, et al. Building watson: An overview of the deepqa project. *AI magazine*, 31(3):59–79, 2010.

Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical methods for rates and proportions*. John Wiley & Sons, 2013.

Mary Ellen Foster, Andre Gaschler, Manuel Giuliani, Amy Isard, Maria Pateraki, and Ronald Petrick. Two people walk into a bar: Dynamic multi-party social interaction with a robot agent. In *Proceedings of the 14th ACM international conference on Multimodal interaction*, pages 3–10. ACM, 2012.

Shinya Fujie, Kenta Fukushima, and Tetsunori Kobayashi. A conversation robot with back-channel feedback function based on linguistic and nonlinguistic information. In *Proc. Int. Conference on Autonomous Robots and Agents*, pages 379–384, 2004.

Shinya Fujie, Riho Miyake, and Tetsunori Kobayashi. Spoken dialogue system using recognition of user ' s feedback for rhythmic dialogue. In *Proc. of Speech Prosody*, 2006.

Shinya Fujie, Daichi Watanabe, Yuhi Ichikawa, Hikaru Taniyama, Kosuke Hosoya, Yoichi Matsuyama, and Tetsunori Kobayashi. Multi-modal integration for personalized conversation: Towards a humanoid in daily life. In *Humanoid Robots, 2008. Humanoids 2008. 8th IEEE-RAS International Conference on*, pages 617–622. IEEE, 2008.

Masahiro Fujita. Digital creatures for future entertainment robotics. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, volume 1, pages 801–806. IEEE, 2000.

Masahiro Fujita. Aibo: Toward the era of digital creatures. *The International Journal of Robotics Research*, 20(10):781–794, 2001.

Adrian Furnham. Language and personality. 1990.

Daniel Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, 27(12):1775–1787, 2009.

Daniel Gatica-Perez, Iain McCowan, Dong Zhang, and Samy Bengio. Detecting group interest-level in meetings. In *ICASSP (1)*, pages 489–492. Citeseer, 2005.

Albert Gatt and Ehud Reiter. Simplenlg: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 90–93. Association for Computational Linguistics, 2009.

Michael S Gazzaniga. *The social brain: Discovering the networks of the mind*. Basic Books New York, 1985.

James Glass, Giovanni Flammia, David Goodine, Michael Phillips, Joseph Polifroni, Shinsuke Sakai, Stephanie Seneff, and Victor Zue. Multilingual spoken-language understanding in the mit voyager system. *Speech communication*, 17(1):1–18, 1995.

Erving Goffman. On face-work. *Interaction ritual*, pages 5–45, 1967.

Erving Goffman. *Forms of talk*. University of Pennsylvania Press, 1981.

H Paul Grice. Logic and conversation. *1975*, pages 41–58, 1975.

Barbara J Grosz, Scott Weinstein, and Aravind K Joshi. Centering: A framework for modeling the local coherence of discourse. *Computational linguistics*, 21(2):203–225, 1995.

A Paul Hare. Types of roles in small groups a bit of history and a current perspective. *Small Group Research*, 25(3):433–448, 1994.

A Hartholt, D Traum, SC Marsella, A Shapiro, G Stratou, A Leuski, LP Morency, and J Gratch. All together now: Introducing the virtual human toolkit. In *International Conference on Intelligent Virtual Humans (Edinburgh, UK*, 2013.

Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5–53, 2004.

Dirk Heylen, Stefan Kopp, Stacy C Marsella, Catherine Pelachaud, and Hannes Vilhjálmsson. The next step towards a function markup language. In *Intelligent Virtual Agents*, pages 270–280. Springer, 2008.

Ryuichiro Higashinaka, Marilyn A Walker, and Rashmi Prasad. An unsupervised method for learning generation dictionaries for spoken dialogue systems by mining user reviews. *ACM Transactions on Speech and Language Processing (TSLP)*, 4(4):8, 2007.

Masahiko Higashiyama, Kentaro Inui, and Yuji Matsumoto. Acquiring noun polarity knowledge using selectional preferences. In *Proceedings of the 14th Annual Meeting of the Association for Natural Language Processing*, pages 584–587, 2008.

Thomas Hofmann. Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 289–296. Morgan Kaufmann Publishers Inc., 1999.

Michael A Hogg and Scott A Reid. Social identity, self-categorization, and the communication of group norms. *Communication Theory*, 16(1):7–30, 2006.

Janet Holmes. Doing collegiality and keeping control at work: Small talk in government departments. *Small talk*, pages 32–61, 2000.

Katherine Isbister, Hideyuki Nakanishi, Toru Ishida, and Cliff Nass. Helper agent: designing an assistant for human-human interaction in a virtual meeting space. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 57–64. ACM, 2000.

Tatsuzo Ishida, Yoshihiro Kuroki, Jinichi Yamaguchi, Masahiro Fujita, et al. Motion entertainment by a small humanoid robot based on open-r. In *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*, volume 2, pages 1079–1086. IEEE, 2001.

Toru Ishida. Development of a small biped entertainment robot qrio. In *Micro-Nanomechatronics and Human Science, 2004 and The Fourth Symposium Micro-Nanomechatronics for Information-Based Society, 2004. Proceedings of the 2004 International Symposium on*, pages 23–28. IEEE, 2004.

Martin Johansson, Gabriel Skantze, and Joakim Gustafson. Head pose patterns in multiparty human-robot team-building interactions. In *Social Robotics*, pages 351–360. Springer, 2013.

Kristiina Jokinen. Turn taking, utterance density, and gaze patterns as cues to conversational activity. In *Proceedings of ICMI 2011 Workshop Multimodal Corpora for Machine Learning: Taking Stock and Road mapping the Future, Alicante, Spain*, pages 31–36, 2011.

Natasa Jovanović, Anton Nijholt, et al. Addressee identification in face-to-face meetings. Association for Computational Linguistics, 2006.

Natasa Jovanovic, Rieks op den Akker, and Anton Nijholt. A corpus for studying addressing behaviour in multi-party dialogues. *Language Resources and Evaluation*, 40(1):5–23, 2006.

Natasa Jovanović et al. Towards automatic addressee identification in multi-party dialogues. Association for Computational Linguistics, 2004.

Peter H Kahn, Nathan G Freier, Takayuki Kanda, Hiroshi Ishiguro, Jolina H Ruckert, Rachel L Severson, and Shaun K Kane. Design patterns for sociality in human-robot interaction. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 97–104. ACM, 2008.

Jaap Kamps, MJ Marx, Robert J Mokken, and Maarten De Rijke. Using wordnet to measure semantic orientations of adjectives. 2004.

Takayuki Kanda, Hiroshi Ishiguro, and Toru Ishida. Psychological analysis on human-robot interaction. In *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, volume 4, pages 4166–4173. IEEE, 2001.

Takayuki Kanda, Rumi Sato, Naoki Saiwaki, and Hiroshi Ishiguro. A two-month field trial in an elementary school for long-term human–robot interaction. *Robotics, IEEE Transactions on*, 23(5):962–971, 2007.

Sin-Hwa Kang, Jonathan Gratch, Candy Sidner, Ron Artstein, Lixing Huang, and Louis-Philippe Morency. Towards building a virtual counselor: modeling nonverbal behavior during intimate self-disclosure. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 63–70. International Foundation for Autonomous Agents and Multiagent Systems, 2012.

Ichiro Kato. Development of wabot 1. *Biomechanism*, 2:173–214, 1973.

Ichiro Kato, Sadamu Ohteru, Katsuhiko Shirai, Toshiaki Matsushima, Seinosuke Narita, Shigeki Sugano, Tetsunori Kobayashi, and Eizo Fujisawa. The robot musician ‘wabot-2’ (waseda robot-2). *Robotics*, 3 (2):143–155, 1987.

Shohei Kato, Shingo Ohshiro, Hidenori Itoh, and Kenji Kimura. Development of a communication robot ifbot. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 1, pages 697–702. IEEE, 2004.

Michael Katzenmaier, Rainer Stiefelhagen, and Tanja Schultz. Identifying the addressee in human-human-robot interactions based on head pose and speech. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 144–151. ACM, 2004.

Adam Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26:22–63, 1967.

A Kikuchi. Notes on the researches using kiss-18. *Bulletin of the Faculty of Social Welfare, Iwate Prefectural University*, 6(2):41–51, 2004.

Norihide Kitaoka, Masashi Takeuchi, Ryota Nishimura, and Seiichi Nakagawa. Response timing detection using prosodic and linguistic information for humanfriendly spoken dialog systems. *Journal of The Japanese Society for Artificial Intellignece*, 20(3):220–228, 2005.

Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi, and Toshikazu Fukushima. Collecting evaluative expressions for opinion extraction. In *Natural Language Processing–IJCNLP 2004*, pages 596–605. Springer, 2005.

Tetsunori Kobayashi and Shinya Fujie. Conversational robots: An approach to conversation protocol issues that utilizes the paralinguistic information available in a robot-human setting. *Acoustical Science and Technology*, 34(2):64–72, 2013.

Tetsunori Kobayashi, Yoichi Matsuyama, and Shinya Fujie. Strategies and protocols for enjoyable conversation robots. *ICASSP2015*, 2014.

Stefan Kopp, Paul Tepper, and Justine Cassell. Towards integrated microplanning of language and iconic gesture for multimodal output. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 97–104. ACM, 2004.

Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R Thórisson, and Hannes Vilhjálmsson. Towards a common framework for multimodal generation: The behavior markup language. In *Intelligent virtual agents*, pages 205–217. Springer, 2006.

Rohit Kumar and Carolyn P Rosé. Engaging learning groups using social interaction strategies. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 677–680. Association for Computational Linguistics, 2010a.

Rohit Kumar and Carolyn P Rosé. Conversational tutors with rich interactive behaviors that support collaborative learning. In *Workshop on Opportunities for intelligent and adaptive behavior in collaborative learning systems*, page 17, 2010b.

Rohit Kumar and Carolyn P Rosé. Comparing triggering policies for social behaviors. In *Proceedings of the SIGDIAL 2011 Conference*, pages 227–238. Association for Computational Linguistics, 2011.

Rohit Kumar, Hua Ai, Jack L Beuth, and Carolyn P Rosé. Socially capable conversational tutors can be effective in collaborative learning situations. In *Intelligent Tutoring Systems*, pages 156–164. Springer, 2010.

Rohit Kumar, Jack L Beuth, and Carolyn P Rosé. Conversational strategies that support idea generation productivity. In *in Groups, 9 th Intl. Conf. on Computer Supported Collaborative Learning, Hong Kong 160 and Rosé, 2010a) Rohit Kumar, Carolyn P. Rosé, 2010, Conversational Tutors with Rich Interactive Behaviors that support Collaborative Learning, Workshop on Opportunity*. Citeseer, 2011.

Benoit Lavoie and Owen Rambow. A fast and portable realizer for text generation systems. In *Proceedings of the fifth conference on Applied natural language processing*, pages 265–268. Association for Computational Linguistics, 1997.

Jina Lee, David DeVault, Stacy Marsella, and David Traum. Thoughts on fml: Behavior generation in the virtual human communication architecture. *Proceedings of FML*, 2008.

Min Kyung Lee, Jodi Forlizzi, Paul E Rybski, Frederick Crabbe, Wayne Chung, Josh Finkle, Eric Glaser, and Sara Kiesler. The snackbot: documenting the design of a robot for long-term human-robot interaction. In *Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on*, pages 7–14. IEEE, 2009.

Stephen C Levinson. *Putting linguistics on a proper footing: Explorations in Goffman's concepts of participation.* Northeastern University Press, 1988.

François Mairesse and Marilyn Walker. Personage: Personality generation for dialogue. In *Annual Meeting-Association For Computational Linguistics*, volume 45, page 496, 2007.

Bronislaw Malinowski. The problem of meaning in primitive languages. *Language and literacy in social practice: A reader*, pages 1–10, 1994.

Matthew Marge and Alexander I Rudnicky. Towards overcoming miscommunication in situated dialogue by asking questions. In *AAAI Fall Symposium Series-Building Representations of Common Ground with Intelligent Agents, Washington, DC*, volume 3, pages 2–1, 2011.

James R Martin and Peter RR White. *The language of evaluation*. Palgrave Macmillan Basingstoke and New York, 2005.

Yosuke Matsusaka, Tojo Tsuyoshi, and Tetsunori Kobayashi. Conversation robot participating in group conversation. *IEICE transactions on information and systems*, 86(1):26–36, 2003.

Yoichi Matsuyama, Hikaru Taniyama, Shinya Fujie, and Tetsunori Kobayashi. Designing communication activation system in group communication. In *Humanoid Robots, 2008. Humanoids 2008. 8th IEEE-RAS International Conference on*, pages 629–634. IEEE, 2008.

Yoichi Matsuyama, Kosuke Hosoya, Hikaru Taniyama, Hiroki Tsuboi, Shinya Fujie, and Tetsunori Kobayashi. Schema: multi-party interaction-oriented humanoid robot. In *ACM SIGGRAPH ASIA 2009 Art Gallery & Emerging Technologies: Adaptation*, pages 82–82. ACM, 2009.

Yoichi Matsuyama, Shinya Fujie, Hikaru Taniyama, and Tetsunori Kobayashi. Psychological evaluation of a group communication activation robot in a party game. In *Eleventh Annual Conference of the International Speech Communication Association*, pages 3046–3049, 2010.

Yoichi Matsuyama, Yushi Xu, Akihiro Saito, Shinya Fujie, and Tetsunori Kobayashi. Multiparty conversation facilitation strategy using combination of question answering and spontaneous utterances. In *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*, pages 103–111. Springer, 2011.

Yoichi Matsuyama, Akihiro Saito, Shinya Fujie, and Tetsunori Kobayashi. Automatic expressive opinion sentence generation for enjoyable conversational systems. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 00(0):00–00, 2014.

Iain Matthews and Simon Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

Marshall McLuhan. *Understanding media: The extensions of man*. MIT press, 1994.

Sean M McNee, John Riedl, and Joseph A Konstan. Making recommendations better: an analytic model for human-recommender interaction. In *CHI'06 extended abstracts on Human factors in computing systems*, pages 1103–1108. ACM, 2006a.

Sean M McNee, John Riedl, and Joseph A Konstan. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*, pages 1097–1101. ACM, 2006b.

Teruhisa Misu and Tatsuya Kawahara. Speech-based interactive information guidance system using question-answering technique. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 4, pages IV–145. IEEE, 2007.

Hiroyasu Miwa, Tetsuya Okuchi, Hideaki Takanobu, and Atsuo Takanishi. Development of a new human-like head robot we-4. In *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, volume 3, pages 2443–2448. IEEE, 2002.

Tomoko Murakami, Koichiro Mori, and Ryohei Orihara. Metrics for evaluating the serendipity of recommendation lists. In *New Frontiers in Artificial Intelligence*, pages 40–46. Springer, 2008.

Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 61–68. ACM, 2009.

Bilge Mutlu, Takayuki Kanda, Jodi Forlizzi, Jessica Hodgins, and Hiroshi Ishiguro. Conversational gaze mechanisms for humanlike robots. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 1(2):12, 2012.

Tetsuji Nakagawa, Takuya Kawada, Kentaro Inui, and Sadao Kurohashi. Extracting subjective and objective evaluative expressions from the web. In *Universal Communication, 2008. ISUC'08. Second International Symposium on*, pages 251–258. IEEE, 2008.

Tetsuji Nakagawa, Kentaro Inui, and Sadao Kurohashi. Dependency tree-based sentiment classification using crfs with hidden variables. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 786–794. Association for Computational Linguistics, 2010.

Teppei Nakano, Shinya Fujie, and Tetsunori Kobayashi. Monea: message-oriented networked-robot architecture. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 194–199. IEEE, 2006.

Tetsuya Nasukawa and Jeonghee Yi. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, pages 70–77. ACM, 2003.

Allen Newell. *Unified theories of cognition*, volume 187. Harvard University Press, 1994.

Yohei Noda, Yoji Kiyota, and Hiroshi Nakagawa. Discovering serendipitous information from wikipedia by using its network structure. In *ICWSM*, 2010.

Yu Ogura, Hiroyuki Aikawa, Kazushi Shimomura, Akitoshi Morishima, Hun-ok Lim, and Atsuo Takanishi. Development of a new humanoid robot wabian-2. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 76–81. IEEE, 2006.

Junichi Osada, Shinichi Ohnaka, and Miki Sato. The scenario and design process of childcare robot, papero. In *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, page 80. ACM, 2006.

Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics, 2002.

Alexandros Papangelis, Ran Zhao, and Justine Cassell. Towards a computational architecture of dyadic rapport management for virtual agents. In *Intelligent Virtual Agents*, pages 320–324. Springer, 2014.

Fabio Pianesi, Massimo Zancanaro, Elena Not, Chiara Leonardi, Vera Falcon, and Bruno Lepri. Multimodal support to group dynamics. *Personal and Ubiquitous Computing*, 12(3):181–195, 2008.

David Premack and Guy Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(04):515–526, 1978.

Antoine Raux and Maxine Eskenazi. A finite-state turn-taking model for spoken dialog systems. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 629–637. Association for Computational Linguistics, 2009.

Antoine Raux, Brian Langner, Dan Bohus, Alan W Black, and Maxine Eskenazi. Let's go public! taking a spoken dialog system to the real world. In *in Proc. of Interspeech 2005*. Citeseer, 2005.

Brian Ravenet, Angelo Cafaro, Magalie Ochs, and Catherine Pelachaud. Interpersonal attitude of a speaking agent in simulated group conversations. In *Intelligent Virtual Agents*, pages 345–349. Springer, 2014.

Ehud Reiter, Robert Dale, and Zhiwei Feng. *Building natural language generation systems*, volume 33. MIT Press, 2000.

Charles Rich, Brett Ponsler, Aaron Holroyd, and Candace L Sidner. Recognizing engagement in human-robot interaction. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 375–382. IEEE, 2010.

Alexander I Rudnicky, Aasish Pappu, Peng Li, Matthew Marge, and Benjamin Frisch. Instruction taking in the teamtalk system. In *Proceedings of the AAAI Fall Symposium on Dialog with Robots*, volume 4, page 2, 2010.

Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, pages 696–735, 1974.

Harvey Sacks, Gail Jefferson, and Emanuel A Schegloff. *Lectures on conversation*, volume 1. Blackwell Oxford, 1992.

Abran J Salazar. An analysis of the development and evolution of roles in the small group. *Small Group Research*, 27(4):475–503, 1996.

Emanuel A Schegloff. *Sequence organization in interaction: Volume 1: A primer in conversation analysis*, volume 1. Cambridge University Press, 2007.

Emanuel A Schegloff and Harvey Sacks. Opening up closings. *Semiotica*, 8(4):289–327, 1973.

Klaus P Schneider. *Small talk: Analysing phatic discourse*, volume 1. Hitzeroth Marburg, 1988.

Fei Sha and Fernando Pereira. Shallow parsing with conditional random fields. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 134–141. Association for Computational Linguistics, 2003.

D Shibata and T Kobayashi. A study on bundle method in one pass trigram decoder. In *Proc. Autumn Meeting of the ASJ*, pages 151–152, 2001.

Takahiro Shimizu and Takashi Iba. A pattern language for facilitation in experiential learning. *IPSJ SIG Technical Reports*, 2006(29):89–92, 2006.

Candace L Sidner, Cory D Kidd, Christopher Lee, and Neal Lesh. Where to look: a study of human-robot engagement. In *Proceedings of the 9th international conference on Intelligent user interfaces*, pages 78–84. ACM, 2004.

Trey Smith and Reid Simmons. Point-based pomdp algorithms: Improved analysis and implementation. *arXiv preprint arXiv:1207.1412*, 2012.

Helen Spencer-Oatey. (im) politeness, face and perceptions of rapport: unpackaging their bases and interrelationships, 2005.

Amanda Stent and Martin Molina. Evaluating automatic extraction of rules for sentence plan construction. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 290–297. Association for Computational Linguistics, 2009.

Amanda Stent, Rashmi Prasad, and Marilyn Walker. Trainable sentence planning for complex information presentation in spoken dialog systems. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 79. Association for Computational Linguistics, 2004.

Matthew Stone. Lexicalized grammar 101. In *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1*, pages 77–84. Association for Computational Linguistics, 2002.

Matthew Stone, Christine Doran, Bonnie Webber, Tonia Bleam, and Martha Palmer. Microplanning with communicative intentions: The spud system. *Computational Intelligence*, 19(4):311–381, 2003.

Dalmas A Taylor and Irwin Altman. Communication in interpersonal relationships: Social penetration processes. 1987.

Deborah J Terry, Michael A Hogg, and Katherine M White. The theory of planned behaviour: self-identity, social identity and group norms. *British Journal of Social Psychology*, 38(3):225–244, 1999.

Andrea L Thomaz and Crystal Chao. Turn-taking based on information flow for fluent human-robot interaction. *AI Magazine*, 32(4):53–63, 2011.

Andrea Lockerd Thomaz, Cynthia Breazeal, Andrew G Barto, and Rosalind Picard. Socially guided machine learning. *Computer Science Department Faculty Publication Series*, page 183, 2006.

Greg Trafton, Laura Hiatt, Anthony Harrison, Frank Tamborello, Sangeet Khemlani, and Alan Schultz. Act-r/e: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction*, 2(1):30–55, 2013.

J Gregory Trafton, Magda D Bugajska, Benjamin R Fransen, and Raj M Ratwani. Integrating vision and audition within a cognitive architecture to track conversations. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 201–208. ACM, 2008.

J Gregory Trafton, Benjamin Fransen, Anthony M Harrison, and Magdalena Bugajska. An embodied model of infant gaze-following. Technical report, DTIC Document, 2009.

David Traum and Jeff Rickel. Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*, pages 766–773. ACM, 2002.

Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

Wataru Uchida, Chiaki Morita, and Takeshi Yoshimura. Knowledge q&a: Direct answers to natural questions. *NTT DOCOMO Technical Journal*, 14:4–9, 2013.

Herwin van Welbergen, Ramin Yaghoubzadeh, and Stefan Kopp. Asaprealizer 2.0: The next steps in fluent behavior realization for ecas. In *Intelligent Virtual Agents*, pages 449–462. Springer, 2014.

Hannes Vilhjálmsson, Nathan Cantelmo, Justine Cassell, Nicolas E Chafai, Michael Kipp, Stefan Kopp, Maurizio Mancini, Stacy Marsella, Andrew N Marshall, Catherine Pelachaud, et al. The behavior markup language: Recent developments and challenges. In *Intelligent virtual agents*, pages 99–111. Springer, 2007.

Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.

Kazuyoshi Wada and Takanori Shibata. Robot therapy in a care house-its sociopsychological and physiological effects on the residents. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 3966–3971. IEEE, 2006.

Marilyn A Walker, Aravind Aravind Krishna Joshi, and Ellen Ellen Friedman Prince. *Centering theory in discourse*. Oxford University Press, 1998.

Marilyn A Walker, Amanda Stent, François Mairesse, and Rashmi Prasad. Individual and domain adaptation in sentence planning for dialogue. *J. Artif. Intell. Res.(JAIR)*, 30:413–456, 2007.

Nigel Ward and Wataru Tsukahara. A study in responsiveness in spoken dialog. *International Journal of Human-Computer Studies*, 59(5):603–630, 2003.

Richard T Watson, Gerardine DeSanctis, and Marshall Scott Poole. Using a gdss to facilitate group consensus: some intended and unintended consequences. *MIS Quarterly*, pages 463–478, 1988.

Joseph Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

Steve Whittaker and Phil Stenton. Cues and control in expert-client dialogues. In *Proceedings of the 26th annual meeting on Association for Computational Linguistics*, ACL '88, pages 123–130, Stroudsburg, PA, USA, 1988. Association for Computational Linguistics.

Jason Williams and Steve Young. Partially observable markov decision processes for spoken dialog systems. *Computer Speech & Language*, 21(2):393–422, 2007.

Fujio Yoshida and Hiromichi Hori. Shinri-sokutei-shakudoshu (psychological tests) ii, 2001.

Steve Young, Milica Gašić, Simon Keizer, François Mairesse, Jost Schatzmann, Blaise Thomson, and Kai Yu. The hidden information state model: a practical framework for pomdp-based spoken dialogue management. *Computer Speech & Language*, 24(2):150–174, 2010.

Massimo Zancanaro, Bruno Lepri, and Fabio Pianesi. Automatic detection of group functional roles in face to face interactions. In *Proceedings of the 8th international conference on Multimodal interfaces*, pages 28–34. ACM, 2006.

Ran Zhao, Alexandros Papangelis, and Justine Cassell. Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *Intelligent Virtual Agents*, pages 514–527. Springer, 2014.

Jun Zheng, Xiang Yuan, and Yam San Chee. Designing multiparty interaction support in elva, an embodied tour guide. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 929–936. ACM, 2005.

Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*, pages 22–32. ACM, 2005.

Publications

JOURNAL PAPERS

2014 Yoichi Matsuyama, Akihiro Saito, Shinya Fujie and Tetsunori Kobayashi, Automatic Expressive Opinion Sentence Generation for Enjoyable Conversational Systems, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014. (DOI:10.1109/TASLP.2014.2363589)

2014 Yoichi Matsuyama, Iwao Akiba, Shinya Fujie and Tetsunori Kobayashi, Four-Participant Group Conversation: A Facilitation Robot Controlling Engagement Density As the Fourth Participant, Journal of Computer Speech and Language, 2014. (DOI:10.1016/j.csl.2014.12.001)

2012 Shinya Fujie, Yoichi Matsuyama, Hikaru Taniyama, and Tetsunori Kobayashi, Conversation Robot Participating in and Promoting Human-Human Communication, The Transactions of The Institute of Electronics, Information and Communication Engineering (IEICE) A, Vol.J95-A No.1, pp37-45, 2012.

INTERNATIONAL CONFERENCE PAPERS (PEER REVIEWED)

2015 Yoichi Matsuyama, Tetsunori Kobayashi, Towards a Computational Model of Small Group Facilitation, AAAI 2015 Spring Symposia Turn-taking and Coordination in Human-Machine Interaction, March 2015.

2013 Yoichi Matsuyama, Iwao Akiba, Akihiro Saito and Tetsunori Kobayashi, A Four-Participant Group Facilitation Framework for Conversational Robots, Association for Computational Linguistics, Proceedings of the SIGDIAL 2013 Conference, Metz, France, pp.284-293, August 2013.

2011 Yoichi Matsuyama, Yushi Xu, Akihiro Saito, Shinya Fujie and Tetsunori Kobayashi, Multiparty Conversation Facilitation Strategy Using Combination of Question Answering and Spontaneous Utterances, IWSDS 2011 Workshop on Paralinguistic Information and its Integration in Spoken Dialogue Systems, pp.99-107, September 2011.

2010 Yoichi Matsuyama, Shinya Fujie, Hikaru Taniyama and Tetsunori Kobayashi, Framework of Communication Activation Robot Participating in Multiparty Conversation, AAAI 2010 Fall Symposia Dialog with Robots, pp.68-73, November 2010.

2010 Yoichi Matsuyama, Shinya Fujie, Hikaru Taniyama and Tetsunori Kobayashi, Psychological Evaluation of A Group Communication Activation Robot in A Party Game, Proceedings of Interspeech 2010, pp.3046-3049, September 2010.

2009 Yoichi Matsuyama, Kosuke Hosoya, Hikaru Taniyama, Hiroki Tsuboi, Shinya Fujie, Tetsunori Kobayashi, SCHEMA: multi-party interaction-oriented humanoid robot, ACM SIGGRAPH ASIA 2009 Art Gallery & Emerging Technologies: Adaptation, pp. 82-82, December 2009.

2009 Shinya Fujie, Yoichi Matsuyama, Hikaru Taniyama, and Tetsunori Kobayashi, Conversation robot participating in and activating a group communication, Proceedings of Interspeech 2009, pp.264-267, September 2009

2009 Yoichi Matsuyama, Hikaru Taniyama, Shinya Fujie, Tetsunori Kobayashi, System Design of Group Communication Activator: An Entertainment Task for Elderly Care, ACM/IEEE Human-Robot Interaction 2009, San Diego, pp.243-244, March 2009.

2008 Yoichi Matsuyama, Hikaru Taniyama, Shinya Fujie, and Tetsunori Kobayashi, Designing Communication Activation System in Group Communication, Proceedings of Humanoids 2008, pp.629-634, December 2008.

2008 Shinya Fujie, Daichi Watanabe, Yuhi Ichikawa, Hikaru Taniyama, Kosuke Hosoya, Yoichi Matsuyama, and Tetsunori Kobayashi, Multi-modal Integration for Personalized Conversation: Towards a Humanoid in Daily Life, Proceedings of Humanoids 2008, pp.617-622, December 2008.

INVITED PAPERS

2014 Yoichi Matsuyama, Jun Nakagawa, Taiki Watai, Akihiro Hayashi, Atsushi Enta and Yasutaka Wada, Designing Human Behaviors: Human-Environment Interaction Design Implicitly Triggering Behavior Changes, IPSJ Magazine Vol.55, No.9, pp.952-954, Information Processing Society of Japan, September 2014.

WORKSHOPS AND TALKS

2014 Yoichi Matsuyama, Embodied Language on Humanoid Robots, Symposium on Robots as Media, School of Culture, Media and Society, Waseda University, December 2014.

2013 Iwao Akiba, Yoichi Matsuyama and Tetsunori Kobayashi, A Facilitation Robot Harmonizing A Four-Participant Conversation, International Workshop on Language and Speech Science, October 2013.

2013 Yoichi Matsuyama, SCHEMA: A Framework of Embodied Conversational Robots Facilitating Small Groups, Embodied Situated & Language Processing, Potsdam, July 2013.

2011 Yoichi Matsuyama, Conversation Robot Participating in and Promoting Multiparty Conversation, Workshop on Social Robots For Assisted Living, University of Aalborg, Denmark, November 2011.

2011 Yoichi Matsuyama, Multiparty Conversation Facilitation Strategies, The 8th Global COE International Symposium on Ambient SoC, July 2011.

2009 Yoichi Matsuyama, Shinya Fujie, Hikaru Taniyama and Tetsunori Kobayashi, Evaluation of Group Communication Activation Robot, International Workshop on Language and Speech Science, October 2009.

2009 Yoichi Matsuyama, Evaluation of Group Communication Activation System, The 5th Global COE International Symposium on Ambient SoC, September 2009. (Outstanding Academic Achievements and Exceptional Performance)

2008 Yoichi Matsuyama, Shinya Fujie, Hikaru Taniyama and Tetsunori Kobayashi, Designing Communication Activation System in Group Communication, International Workshop on Language and Speech Science, September 2008.

DOMESTIC CONFERENCE PAPERS

2014 Yoichi Matsuyama, Alexandros Papangelis, Ran Zhao and Justine Cassell, Dyadic Computational Model of Rapport Management, The Japanese Society for Artificial Intelligence (JSAI), SIG-SLUD, December 2014.

2013 Yoichi Matsuyama, Akihiro Saito and Tetsunori Kobayashi, Automatic Opinion Generation for Serendipitous Question Answering Systems, Acoustical Society of Japan (ASJ) 2013 Autumn Meeting, NO.3-8-3, September 2013.

2013 Iwao Akiba, Yoichi Matsuyama and Tetsunori Kobayashi, Facilitation Strategies for A Robot Maintaining Four-Participant Groups, Acoustical Society of Japan (ASJ) 2013 Autumn Meeting, NO.3-8-2, September 2013.

2013 Iwao Akiba, Yoichi Matsuyama and Tetsunori Kobayashi, Procedures of Obtaining Initiatives for Multiparty Conversation Facilitation Robots, Information Processing Society of Japan (IPSJ), SIG-SLP 97(10), 1-8, July 2013.

2013 Yoichi Matsuyama, Akihiro Saito, Atsushi Ito, Iwao Akiba, Moemi Watanabe and Tetsunori Kobayashi, Active Timing Detection and Strategies for Multiparty Conversation Facilitation Systems, The Japanese Society for Artificial Intelligence (JSAI), SIG-SLUD-B203-05, pp.17-24, February 2013 (Annual best presentation, The Japanese Society for Artificial Intelligence SIG-SLUD).

2013 Akihiro Saito, Yoichi Matsuyama, Azusa Todoroki and Tetsunori Kobayashi, Natural Sentence Generation for Serendipitous Question Answering Systems, The Japanese Society for Artificial Intelligence (JSAI), SIG-SLUD-B203-01, pp.1-6, February 2013.

2012 Yoichi Matsuyama, Akihiro Saito, Iwao Akiba, Moemi Watanabe and Tetsunori Kobayashi, Facilitation Robot Promoting the Greatest Participation of the Greatest Number in Multiparty Conversation, Human-Agent Interaction Symposium 2012, 2B-3, December 2012 (HAI-2012 Outstanding Research Award).

2011 Yoichi Matsuyama, Shinya Fujie, Akihiro Saito and Tetsunori Kobayashi, Patterns of Strategies for Multiparty Conversation Facilitation Systems, Annual Conference of the Japanese Society for Artificial Intelligence (JSAI), 1O2-OS18-9, June 2012.

2011 Akihiro Saito, Yoichi Matsuyama, Shinya Fujie and Tetsunori Kobayashi, Evaluation of Multiparty Conversation Facilitation Strategies of A Conversational Robot, Technical Report of The Institute of Electronics,Information and Communication Engineering (IEICE), vol.111, no.225, SP2011-53, pp. 7-12, October 2011.

2011 Shinya Fujie, Yoichi Matsuyama, Akihiro Saito and Tetsunori Kobayashi, Development of conversation robot participating in multiparty conversation and promoting communication, Acoustical Society of Japan (ASJ) 2011 Autumn Meeting, NO.3-10-6, September 2011.

2010 Akihiro Saito, Yoichi Matsuyama, Shinya Fujie and Tetsunori Kobayashi, Multiparty Conversation Facilitation Strategy Using Combination of Question Answering and Spontaneous Utterance, Annual Conference of the Japanese Society for Artificial Intelligence (JSAI), 3C2-OS19-10, June 2011.

2010 Yoichi Matsuyama, Shinya Fujie, Akihiro Saito, Xu Yushi and Tetsunori Kobayashi, Communication Activation Oriented Conversation Robot: An Application in an Elderly Care Facility, Technical Report of The Institute of Electronics, Information and Communication Engineering (IEICE), Vol.10, No.219, pp.7-12, October 2010.

2010 Shinya Fujie, Yoichi Matsuyama and Tetsunori Kobayashi, Group Communication Activation Robot, The Japanese Society for Artificial Intelligence (JSAI), SIG-SLUD-B002, pp.7-10, October 2010.

2010 Yoichi Matsuyama, Shinya Fujie, Hikaru Taniyama and Tetsunori Kobayashi, Impression Evaluation of Group Communication Activation Robot, The Japanese Society for Artificial Intelligence (JSAI), SIG-SLUD-B001-02, pp.7-12, July 2010.

2010 Hikaru Taniyama, Yoichi Matsuyama, Shinya Fujie and Tetsunori Kobayashi, Development of Group Communication Robot based on Behavior Design with Participation Structure, The Japanese Society for Artificial Intelligence (JSAI), SIG-SLUD-A903, pp.55-60, February 2010.

2009 Hikaru Taniyama, Yoichi Matsuyama, Shinya Fujie and Tetsunori Kobayashi, Behavior Analysis of Group Communication Robot Based on Participation Structure, The Japanese Society for Artificial Intelligence (JSAI), SIG-SLUD-A901, pp.1-6, July 2009.

2008 Yoichi Matsuyama, Shinya Fujie, Hikaru Taniyama and Tetsunori Kobayashi, Communication Activation System in Group Communication, The Japanese Society for Artificial Intelligence (JSAI), SIG-SLUD-A801, pp.15-22, July 2008 (Annual best presentation, The Japanese Society for Artificial Intelligence SIG-SLUD).

PATENTS

2010 Conversational Robot (Japan 2010-221556)

2008 Conversational Facilitation System and Robot (Japan 2008-304140)