Journal of Information Processing Vol.32 35–40 (Jan. 2024)

[DOI: 10.2197/ipsjjip.32.35]

Technical Note

Evaluation of Prosodic Features Suitable for Conversational Agents Replying with a Joke

Kenro Go[1] Shunsuke Tokuda[2] Haruna Niiyama[2]
Toshiki Onishi[1] Asahi Ogushi[1] Akihiro Miyata[2,a)]

Received: April 10, 2023, Accepted: July 5, 2023

Abstract: Conversational agents are becoming popular in various daily situations. To make agents more user-friendly in the future, we are studying ways to make them more humorous. Our proposed interaction style for the agents is to perform boke (performing jokes) and tsukkomi (butting to jokes) based on Japanese manzai during communication with the user. To improve this style, we focus on the prosodic features of the agent's voice which affect the user's impression of a agent. The results of the experiment suggest that the user's sense of humor, friendliness, and motivation to continue the conversation may be improved by making the agent's voice speed normal and the pitch high. Also, it was suggested that the faster speed for the agent's voice may not be suitable for improving the user's sense of humor, familiarity, and motivation to continue the conversation.

Keywords: Humor, Joke, Manzai, Conversational agent, Human–agent interaction, prosodic

1. Introduction

The number of cases in which AI has been introduced into voice input systems is increasing [1]. In particular, the number of AI assistants that rely on voice input, such as Siri [2] and Google Assistant [3], is rapidly increasing in various everyday scenarios [4]. In the future, conversational agents are expected to be widely used in situations where verbal communication with users is important, such as chat partner and caregiver. For such technologies to perform a more important role, it is necessary for people to feel familiar *1 with automated agents. To achieve this, we focused on humor, which enables automated agents to establish relationships with humans [5], [6], [7]. Some studies aimed to enable users to become familiar with an agent by including humor in its behavior. We also proposed an interaction style for agents to create humorous statements during conversations with users in real time, referring to Japanese Manzai *2 [8], [9], [10], [11], [12]. However, previous research focused only on the linguistic features of humorous statements and did not consider nonverbal features. The use of conversational agents that rely on voice-based inputs and outputs has rapidly increased in people's daily lives. Therefore, clarifying nonverbal features that enhance the effects of humor in these agents is important.

The use of voice as a modality for interaction between conversational agents and users is becoming increasingly indispensable [13]. Thus, we focused on prosodic features among the various nonverbal features that perceptibly affect humor. Based on the aforementioned factors, this study aims to clarify the prosodic features of voice that are suitable for agents' humorous statements. This paper describes a controlled experiment in which a conversational agent delivered humorous statements during a conversation with a user at different speeds and pitches, and analyzes the change in the user's impression of the conversational agent's humorous expressions.

2. Related Work

2.1 Effects of Humor in Communication

Humor has been studied from various academic perspectives such as philosophy, psychology, and linguistics [14]. Humor is effective in many ways by helping people to feel good about themselves. For example, humor is considered to have a positive effect on the human spirit. Humor is considered important in forming friendly relationships between people [5]. Holmes [15] examined the effects of humor on physical and mental health in the workplace and on self-evaluation of one's job performance and found that humor exhibited positive effects in these areas. Humor is effective in forming relationships between people and automated agents [6], [7], [16].

2.2 Agent-generated Humor

Two types of humorous expressions can be applied to automatic conversational agents: scenario-based humor [17], [18], [19] and situational humor production in response to user's statements [20], [21], [22].

As scenario-based humor, Haraguchi et al. [17] created two

1 Graduate School of Integrated Basic Sciences, Nihon University, Setagaya, Tokyo 156–8550, Japan.
2 College of Humanities and Sciences, Nihon University, Setagaya, Tokyo 156–8550, Japan.
a) miyata.akihiro@acm.org

*1 In this study, familiarity represents friendliness and intimacy.
*2 Manzai is a traditional Japanese comedic style. In Manzai, two persons perform humor by playing roles of boke (performing jokes) and tsukkomi (butting to jokes).

agents designed to automatically generate the *Manzai* from news articles. First, the system analyzes the sentences in a Web news article for emotion. Then, the agents generated humor by having the News article read with the opposite facial expression to that of the analysis results. Lee et al. [18] proposed a method to add humorous titles to images. In their study, images and one or more humorous titles were used as the datasets for machine learning. This enables the system to consider the regions, trends, and other factors to improve the ability to generate humor. Umetani et al. [19] proposed a method for robots to perform *Manzai*. The *Manzai* script was automatically generated from online news articles.

As an example of situational humor, Hans and Anton [20] used the relationship between pronouns to generate humorous sentences. Humor is generated by finding a target word that differs from the correct target word by referring to a pronoun. Oliviero and Carlo [21] created humor using short names, expressed by connecting the initial letter of each word in a sentence and replacing the component words without changing the abbreviation. When generating humorous sentences, words were obtained from WordNet [23] to find a word with pronunciation as close to that of the original word with an ironic meaning as possible. Dybala [22] proposed an agent designed to generate a pun using words in user statements with different usages. For example, when a user says "Kaeru daikirai!" ("I hate frogs" in English), the agent might reply with "Kaeru to ieba tsukaeru no desu ne" ("Speaking of frogs, we could use one!" in English).

2.3 Phonetic Expression

Many studies were conducted to analyze the way different prosodic features affect people. Shibata et al. [24] found that participants preferred a higher-pitched robot while conversing with both higher- and lower-pitched robots. Lee et al. [25] discovered that male voices are perceived as more convincing, socially appealing, and reliable than female voices, thereby supporting gender stereotypes. Nass et al. [26] discovered that synthetic voices with masculine characteristics are perceived as friendlier and more competent than those with feminine characteristics. Additionally, they observed that the gender of the voice was better received when it matched the stereotypical expectations of gendered speech content. Read and Belpaeme [27] discovered that, although children could recognize emotions through nonverbal expressions, a lack of agreement existed among them regarding which emotions were being conveyed. Furthermore, pitch did not have an impact on the attribution of emotions. Iizuka et al. [28] studied the effects of the synthetic voices of conversational agents trained in natural language. The results of the experiment confirmed that the participants who interacted with a conversational agent with a synthetic voice based on natural language tended to have shorter response times. In addition, they confirmed that the tendency to evaluate interactions with a conversational agent is similar to interactions with a human.

3. Research Goal

The humor of a conversational agent is considered effective for familiarizing users with the agent [5], [6], [7]. However, in the studies in Sections 2.1 and 2.2, the agents could not make humorous statements in real time while conversing with users. Based on this, we proposed an interaction style that references Japanese Manzai, in which the agents deliver humorous statements while conversing with the user in real time [8], [9], [10], [11], [12]. These existing studies focused only on the verbal features in humorous statements of conversational agents and limited the interaction between the user and interactive agent to text. Therefore, we focused on nonverbal features to enhance the effect of humor in the agent's utterances based on previous research. Many nonverbal factors influence the user's impression of an agent, such as voice, movement, and facial expressions. In particular, we focused on voice, which is less affected by the physical standards of conversational agents.

Conversational agents that rely on voice-based input and output have rapidly increased in our daily lives. In addition, the use of voice-based communication as a means for interaction between conversational agents and users has become increasingly essential [13]. Therefore, clarifying of prosodic features that enhance the effects of humor on these agents is important. Section 2.3 suggests that prosodic features have various effects on users. However, previous studies did not focus on the effects of conversational agents on the prosodic features of humorous statements.

Given the above discussion, we set our research goal as to clarifying whether users' impressions of an agent differ depending on the prosodic features of the agent's humorous statement.

4. Implementation

This section describes the implementation of the system used for verification to achieve our research goals. The system consists of the humor generation part and the voice output part.

For humor generation, we used the conversational agent that was proposed in our previous study [8], [9], [10], [11], [12]. In a previous study, a conversational agent performed mishearing humor, in which it replaced a word in the user's statement with another word and listened back to the user. Specifically, the agent selects a word that is distant from the topic of the user's statements, close in pronunciation to the replaced word, and likely to be recognized by the user, and generates mishearing humor statements. The user's input to the system is voice. To recognize the user's voice, we used the SpeechRecognition library in Python.

4.1 Humor Generation Part

The humor generation process entails selecting a replaced word in the user's input and determining the way to mishear the selected word.

All nouns in the user input are considered as candidates to be replaced. Subsequently, for each replaced word candidate, the similarity (u) of the concept to the nouns in the statements, excluding itself, is calculated. The s_t values of each noun in the replaced word candidates are assigned higher scores as the similarity (u) between the words increases. Therefore, s_t is calculated as shown in Eq. (1), where n is the number of nouns in the statements, and $u_{(k)}$ is the similarity of the concept between the k-th word and the candidate word considered for replacement. The noun with the highest s_t in the user input is determined to be the replaced word.

$$s_t = \frac{1}{n-1} \sum_{k=1,k\neq i}^{n} u_{(k)} \tag{1}$$

We generate a corpus by performing a morphological analysis of the full text of Japanese Wikipedia articles using MeCab [29], removing unnecessary parts of speech such as particles without lexical meaning, and separating them. All the words in the corpus are used as candidates for the misheard word. A word that is distant in meaning from the topic of the user's statements, close in pronunciation to the replaced word, and likely to be recognized by the user was selected from these candidates.

To select a word with a distant meaning, a score (s_s) was calculated, which indicates the distance of the meaning using the inverse of the similarity ($u_{(k)}$) between the replaced word and the topic of the user's statements. Thus, s_s is calculated as shown in Eq. (2), where n is the number of nouns in the statement, and $u_{(k)}$ is the similarity of the concept between the k-th word and the candidate word for the misheard word.

$$s_s = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{2 + u_{(k)}} \tag{2}$$

To calculate s_p, the edit distance [*3] (d_e) between the vowels of each misheard word candidate and the replaced word is measured. The closer the distance, the higher the value of s_p. Thus, s_p is calculated as shown in Eq. (3).

$$s_p = \frac{1}{1 + d_e} \tag{3}$$

The score (s_r), which indicates the efficiency of recognizing each word in the misheard word candidate list, is calculated by counting the number of times (hereinafter called f) each word appears in Wikipedia. The number of times a word appears follows a power distribution. Therefore, to prevent words with an extremely high number of appearances from significantly impacting the selection of misheard words, the logarithm of the number of appearances was considered as s_r. Thus, the value of s_r was calculated using Eq. (4):

$$s_r = \log f \tag{4}$$

The s_s, s_p, and s_r values calculated for each misheard word candidate were normalized and weighted, and their total was assigned as the final score (s_h) for each misheard word candidate. The weights for s_s, s_p, and s_r were set to 1.5, 2.0, and 1.0, respectively, according to a discussion among the authors based on the experimental results of previous studies [9], [10], [11], [12]. The misheard word candidate with the highest s_h value was applied to the following template as a misheard word:

Template of misheard statement: "E? [misheard word]?". In English, this template means "What? [misheard word]?".

Examples of the output of the system are shown below.

User: Kirei na *shashin* ga toreta ("I took a beautiful photo.").
Agent: E? *Kashin*? ("What? retainer?")
User: *Aisha* ga kirei de ureshii ("I'm glad my car is clean.").
Agent: E? *Taisya*? ("What? Metabolism?")
User: *Saibankan* ni narutameni benkyo suruzo! ("I study to be a judge.").
Agent: E? *Aianman?* ("What? Ironman?")

4.2 Voice Output Part

We used Open JTalk for voice synthesis [30]. Open JTalk is an HMM-based text-to-speech system that generates speech based on Japanese text. The speed and pitch of voice synthesis are set in reference to Kyoko, which is an Apple voice assistant that boasts a large share in Japan. Specifically, when "ohayou gozaimasu" ("Good morning" in English.) is output, the time from start to end is 1.04 s, and the average fundamental frequency is 224 Hz (hereinafter called normal voice). Previous research has shown that the robot responds to a user's statement by delaying its response by 1.0 s rather than responding immediately [31]. Thus, the agent responds 1.0 s after the user's voice input ends.

5. Experiment

We evaluated the prosodic features of the agent's voice for the humorous statement. Specifically, we verified whether differences in the speeds and pitches of the agent's voice affected the user's sense of humor, familiarity, and motivation to continue interacting with the agent.

5.1 Conditions of the Experiment

Ten people in their twenties participated in this study. Half of the participants used interactive agents, such as Siri and Google Assistant in their daily life. Participants used an agent in a scene that involved talking about a recent memorable event. We used nine conditions in which the speed (normal, fast, and slow) and pitch (normal, high, and low) of the agent's voice differed (**Table 1**). For voice speed, the time from the start to the end of the voice was half the time of normal for the fast condition and 2.0 times longer for the slow condition.

For the voice pitch, the method of generation of high and low pitches relative to the normal pitch was based on the experimental conditions of the existing study [32]. Specifically, the average fundamental frequency of the voice was set to 156.25% at high condition, and 64.0% at low condition.

5.2 Procedure

In this experiment, the participants communicated about a recent memorable event once, each with nine conditions of agents with different voice speeds and pitches. Participants considered

Table 1 Conditions used in the experiment.

Speed	Pitch
Normal	Normal
Fast	Normal
Slow	Normal
Normal	High
Normal	Low
Fast	High
Fast	Low
Slow	High
Slow	Low

*3 edit distance between two strings indicates their difference. This is defined as the minimum number of steps required to transform one string into another by inserting, deleting, or replacing characters

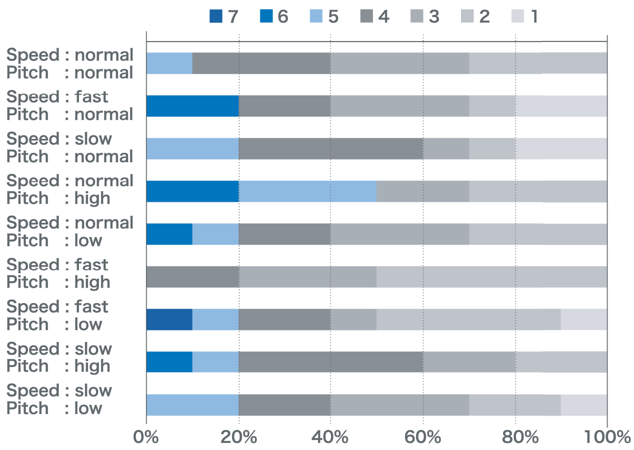

Fig. 1 Degree of humor experienced by the users (N = 10).

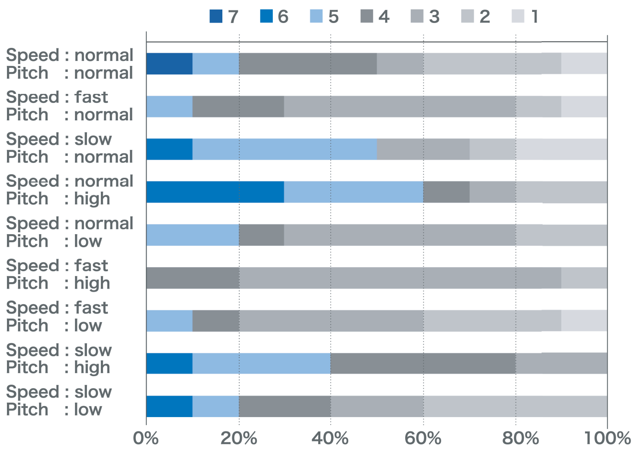

Fig. 2 Degree of humor experienced by the users (N = 10).

nine statements used in each interaction. The participants interacted with the agent using a PC. To avoid any influence on the experiment, the participants wore headsets, and no one except for one experimenter entered the room. After each interaction under each condition, all participants responded to the following questionnaire on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree).

Q1. Did you perceive humor in the statement made by the agent?
Q2. Did you feel familiar with agent?
Q3. Did you feel motivated to continue the conversation with the agent?

5.3 Results and Discussion

Figures 1 to **3** show the results of the questionnaire regarding the degree of humor experienced, familiarity experienced toward the agents, and the user's motivation to continue the conversation with the agent. **Table 2** shows the number of participants who answered five or higher on the Likert scale for Q1–Q3. For each question, we tested several methods for analyzing the results. The results of one-way ANOVA showed no significant differences among the methods. In addition, the results of the two-way ANOVA for the two factors of speed and pitch during voice showed no significant differences among the methods. These results suggest that the user's sense of humor, familiarity, and motivation to continue conversing with the agent were not necessarily affected by differences in the speed or pitch of the voice in the

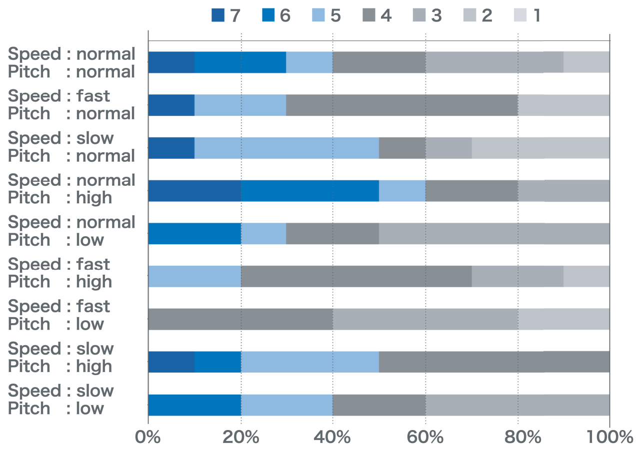

Fig. 3 Degree of the motivation to continue the conversation with the agents (N = 10).

Table 2 The rates of participants who answered 5 and over for Q1 to Q3 (N = 10).

Speed/pitch	Q1: Humor	Q2: Familiarity	Q3: Motivation
Normal/Normal	10%	20%	40%
Fast/Normal	20%	10%	30%
Slow/Normal	20%	50%	50%
Normal/High	50%	60%	60%
Normal/Low	20%	20%	30%
Fast/High	0%	0%	20%
Fast/Low	20%	10%	0%
Slow/High	20%	40%	50%
Slow/Low	20%	20%	40%

agent's humorous statements.

However, in the condition of normal speed and high voice, the rate of participants who answered five or higher in all the questions from Q1 to Q3 was higher than 50%, which was the highest of all conditions. This suggests that the agent's humorous statements at normal speed and in a high voice may improve the user's sense of humor, familiarity with the agent, and motivation to continue the conversation. To discuss the reasons for this result, we interviewed the participants and found that some of them said that the slightly higher voice of the agent was more characterful and friendly. In addition, we confirmed the condition with the lowest rate of responses of five or higher in Q1 to Q3: fast speed and high voice had the lowest rate of humor and friendliness at 0%, and fast speed and low voice had the lowest rate of motivation to continue the conversation at 0%. These results suggest that increasing the voice speed may not be suitable for improving the user's sense of humor, familiarity with the agent, and motivation to continue the conversation. Some participants reported that they had difficulty hearing and were confused by the agent's statement.

6. Conclusion

In this study, we examined the differences in users' perceptions of the agent based on the differences in prosodic features when the agent produced humorous utterances to enhance the interaction with the agent, as proposed in previous studies. Specifically, we conducted an experiment in which users interacted with the agent at different speeds and pitches while performing humor and compared their sense of humor and familiarity with the agent and their motivation to continue interacting with the agent across conditions. The results of the experiment suggest that a user's sense of humor, friendliness, and motivation to continue the conversation can be improved by normalizing the agent's voice speed and increasing the pitch. It was also suggested that a faster speed of the agent's voice may not be suitable for improving the user's sense of humor, familiarity, and motivation to continue conversation.

In the future, we intend to conduct further verification and improve the interaction model in which the agent delivers humorous statements in real time during interactions with users. First, the evaluation was limited to the case of a single agent producing humor statements. In our proposed method, two pairs of agents perform humor with their roles divided between the boke and tsukkomi roles. We will conduct a verification to clarify the appropriate prosodic features of the agents when the tsukkomi role is included. Second, we intend to introduce the timing of humorous statements as a new nonverbal factor in the proposed interaction model. It has been shown that the timing of the presentation of boke and tsukkomi affects the humor of Manzai [33]. However, existing studies have only verified the appropriate timing for one-way Manzai performances, and appropriate timing for performing humor in an interaction scene with a user is not clear. Therefore, we plan to verify the appropriate timing for boke and tsukkomi in response to user statements.

References

[1] Diederich, S., Brendel, A.B. and Kolbe, L.M.: On Conversational Agents in Information Systems Research: Analyzing the Past to Guide Future Work, *Wirtschaftsinformatik* (2019).

[2] Apple Inc.: Siri, available from ⟨https://www.apple.com/ios/siri/⟩.

[3] Google LLC: Google assistant, available from ⟨https://assistant.google.com/⟩.

[4] Knote, R., Janson, A., Söllner, M. and Leimeister, J.M.: Classifying Smart Personal Assistants: An Empirical Cluster Analysis, *Hawaii International Conference on System Sciences* (*HICSS*) (2019).

[5] Khooshabeh, P., McCall, C., Gandhe, S., Gratch, J. and Blascovich, J.: Does It Matter If a Computer Jokes, *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, *CHI EA '11*, pp.77–86, Association for Computing Machinery (2011).

[6] Luger, E. and Sellen, A.: "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents, *Proc. 2016 CHI Conference on Human Factors in Computing Systems*, *CHI '16*, pp.5286–5297, Association for Computing Machinery (2016).

[7] Liao, Q.V., Mas-ud Hussain, M., Chandar, P., Davis, M., Khazaeni, Y., Crasso, M.P., Wang, D., Muller, M., Shami, N.S. and Geyer, W.: All Work and No Play?, *Proc. 2018 CHI Conference on Human Factors in Computing Systems*, *CHI '18*, pp.1–13, Association for Computing Machinery (2018).

[8] Go, K., Nagaoka, D., Nakahara, R., Nakatsuji, M. and Miyata, A.: A Conversational Agent Replying with a joke, *Trans. Virtual Reality Society of Japan*, Vol.23, No.4, pp.231–238 (2018).

[9] Go, K., Nagaoka, D., Nakahara, R. and Miyata, A.: A Conversational Agent Replying with a Humor by Topic-Word Substitution, *Journal of Information Processing Society of Japan*, Vol.61, No.1, pp.113–122 (2020).

[10] Go, K., Tominaga, S., Muto, Y. and Miyata, A.: A Joking System Using Role-Separated Conversational Agents, *Journal of Information Processing Society of Japan*, Vol.61, No.8, pp.1353–1362 (2020).

[11] Go, K., Onishi, T., Ogushi, A. and Miyata, A.: Conversational Agents Replying with a Manzai-Style Joke, *Proc. 33rd Australian Conference on Human-Computer Interaction*, *OzCHI '21*, pp.221–230, Association for Computing Machinery (2022).

[12] Go, K., Onishi, T., Ogushi, A., Tokuda, S. and Miyata, A.: Evaluation of Motivation for Conversational Agents Replying with Manzai-Style Jokes, *Trans. Virtual Reality Society of Japan*, Vol.28, No.1, pp.43–53 (2023).

[13] Seaborn, K., Miyake, N.P., Pennefather, P. and Otake-Matsuura, M.: Voice in Human–Agent Interaction: A Survey, *ACM Comput. Surv.*, Vol.54, No.4, pp.1–43 (2021).

[14] Dynel, M.: Beyond a Joke: Types of Conversational Humour, *Language and Linguistics Compass*, Vol.3, pp.1284–1299 (2009).

[15] Holmes, J.: Sharing a laugh: Pragmatic aspects of humor and gender in the workplace, *Journal of Pragmatics*, Vol.38, No.1, pp.26–50 (2006).

[16] Hampes, W.P.: The relationship between humor and trust, Vol.12, No.3, pp.253–260 (1999).

[17] Haraguchi, K., Yane, K., Sato, A., Aramaki, E., Miyashiro, I. and Nadamoto, A.: Chat-Type Manzai Application: Mobile Daily Comedy Presentations Based on Automatically Generated Manzai Scenarios, *Proc. 18th International Conference on Advances in Mobile Computing & Multimedia*, *MoMM '20*, pp.143–152, Association for Computing Machinery (2021).

[18] Lee, D., Han, S.-H., Oh, K.-J. and Choi, H.-J.: A Temporal Community Contexts Based Funny Joke Generation, *2017 18th IEEE International Conference on Mobile Data Management* (*MDM*), pp.360–365 (2017).

[19] Umetani, T., Aoki, S., Akiyama, K., Mashimo, R., Kitamura, T. and Nadamoto, A.: Scalable Component-Based Manzai Robots as Automated Funny Content Generators, *Journal of Robotics and Mechatronics*, Vol.28, No.6, pp.862–869 (2016).

[20] Hans, T. and Anton, N.: Computational Humour: Utilizing Cross-Reference Ambiguity for Conversational Jokes, Vol.7, No.4578, pp.477–483 (2007).

[21] Oliviero, S. and Carlo, S.: Getting Serious about the Development of Computational Humor, *Proc. 18th International Joint Conference on Artificial Intelligence*, *IJCAI'03*, pp.59–64, Morgan Kaufmann Publishers Inc. (2003).

[22] Dybala, P.: Multiagent system for joke generation: Humor and emotions combined in human-agent conversation, *Thematic Issue on Computational Modeling of Human-Oriented Knowledge within Ambient Intelligence of The IOS Journal on Ambient Intelligence and Smart Environments, to appear in January 2010* (2010).

[23] Miller, G.A.: WordNet: A Lexical Database for English, *Comm. ACM*, Vol.38, No.11, pp.39–41 (1995).

[24] Shibata, R., Fukada, C., Kojima, T., Sato, K., Hashikura, Y., Ozeki, M. and Oka, N.: Does Talking to a Robot in a High-Pitched Voice Create a Good Impression of the Robot?, *2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, pp.19–24 (2012).

[25] Lee, E.J., Nass, C. and Brave, S.: Can Computer-Generated Speech Have Gender? An Experimental Test of Gender Stereotype, *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, *CHI EA '00*, pp.289–290, Association for Computing Machinery (2000).

[26] Nass, C., Moon, Y. and Green, N.: Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices, *Journal of Applied Social Psychology*, Vol.27, pp.864–876 (1997).

[27] Read, R. and Belpaeme, T.: How to Use Non-Linguistic Utterances to Convey Emotion in Child-Robot Interaction, *Proc. 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, *HRI '12*, pp.219–220, Association for Computing Machinery (2012).

[28] Iizuka, T. and Mori, H.: How Does a Spontaneously Speaking Conversational Agent Affect User Behavior?, *IEEE Access*, Vol.10, pp.111042–111051 (2022).

[29] MeCab: Yet another part-of-speech and morphological analyzer, available from ⟨http://taku910.github.io/mecab/⟩.

[30] Nagoya Institute of Technology: Open jtalk, available from ⟨https://open-jtalk.sp.nitech.ac.jp/⟩.

[31] Shiwa, T., Kanda, T., Imai, M., Ishiguro, H. and Hagita, N.: How Quickly Should Communication Robots Respond?, *Proc. 3rd ACM/IEEE International Conference on Human Robot Interaction*, *HRI '08*, pp.153–160, Association for Computing Machinery (2008).

[32] Uchida, T.: Effects of the speech rate on speakers' personality-trait impressions, *Shinrigaku Kenkyu*, Vol.73, No.2, pp.131–139 in Japanese (2002).

[33] Man, Z., Noriko, S. and Mamiko, S.: Analysis of timing structure in Xiangsheng and Manzai: Comparative study of performing arts between China and Japan, *Technical Report of IEICE. HIP*, Vol.113, No.73, pp.167–172 (2013).

Kenro Go is a doctoral program's student in the Graduate School of Integrated Basic Sciences, Nihon University. He received his M.S. degree from the Graduate School of Integrated Basic Sciences, Nihon University in 2020. He is engaged in research on human-computer interaction.

Shunsuke Tokuda is an undergraduate student in the College of Humanities and Sciences, Nihon University.

Haruna Niiyama is an undergraduate student in the College of Humanities and Sciences, Nihon University.

Toshiki Onishi is a doctoral program's student in the Graduate School of Integrated Basic Sciences, Nihon University. He received his M.S. degree from the Graduate School of Integrated Basic Sciences, Nihon University in 2022. He is engaged in research on multimodal interaction and human-computer interaction.

Asahi Ogushi is a graduate student in the Graduate School of Integrated Basic Science, Nihon University. He received his B.S. degree from the science from Nihon University in 2022. He is engaged in research on multimodal interaction and human-computer interaction.

Akihiro Miyata is a professor in Nihon University. He received his Ph.D. degree in engineering in 2008 from Keio University. In 2016, he joined Nihon University as an associate professor. He is a senior member of IPSJ, and a member of ACM and VRSJ.