Analysis of factors influencing the impression of speaker individuality in android robots*

Ryusuke Mikata[1], Carlos T. Ishi[2], Takashi Minato[3] and Hiroshi Ishiguro[4]

***Abstract*— Humans use not only verbal information but also non-verbal information in daily communication. Among the non-verbal information, we have proposed methods for automatically generating hand gestures in android robots, with the purpose of generating natural human-like motion. In this study, we investigate the effects of hand gesture models trained/designed for different speakers on the impression of the individuality through android robots. We consider that it is possible to express individuality in the robot, by creating hand motion that are unique to that individual. Three factors were taken into account: the appearance of the robot, the voice, and the hand motion. Subjective evaluation experiments were conducted by comparing motions generated in two android robots, two speaker voices, and two motion types, to evaluate how each modality affects the impression of the speaker individuality. Evaluation results indicated that all these three factors affect the impression of speaker individuality, while different trends were found depending on whether or not the android is copy of an existent person.**

I. INTRODUCTION

In recent years, the opportunities for robots to play an active role in society have been expanded, so that the opportunities for human-robot interaction have increased. For example, we have come to see a scene where robots that look close to humans like Pepper and Nao, so-called humanoid robots, interact with people through product introductions and directions around the city. In such interactions, not only verbal information, but also nonverbal information exchange is inevitable. However, at present the robot lacks the technology to express such non-verbal information. In the situation of human-to-human conversation, the listener reads not only the information from the speaker's voice but also the movements of the head, body, hands and facial expression. The speaker is conscious of the listener and adjusts his / her speech and body movement. By expressing this non-verbal information in human-robot interaction, it becomes easier to understand what the robot wants to communicate to the other person, leading to the realization of a more natural conversation. In addition, it may lead to the expression of emotions and tensions, and by improving the impression on robots, it may be possible to further expand its activity field.

*This work was supported by JST, ERATO, Grant Number JPMJER1401.

[1]R. Mikata is with the ATR Hiroshi Ishiguro Labs., Kyoto, Japan (e-mail: mikata.ryusuke@irl.sys.es.osaka-u.ac.jp).

[2]C. T. Ishi is with the ATR Hiroshi Ishiguro Labs., Kyoto, Japan (e-mail: carlos@atr.jp).

[3]T. Minato is with the ATR Hiroshi Ishiguro Labs., Kyoto, Japan (e-mail: minato@atr.jp).

[4]H. Ishiguro is with the ATR Hiroshi Ishiguro Labs., Kyoto, Japan (e-mail: ishiguro@sys.es.osaka-u.ac.jp).

In our laboratory, we have been researching methods for the generation of lip, head, facial expression, and hand movements accompanying the robot's speech[1], [2], [3], [4], [5], [6]. Since android robots, such as the one illustrated in Fig. 1, closely resemble a human being, we think that more human-like motion is required than general humanoid robots.

Fig. 1. The android robot ERICA.

Android robots can be a copy of an existing person (also called geminoids), or a newly created human-like entity, such as the android ERICA shown in Fig. 1. In geminoids, the appearance is made to be identical to the target speaker. However, if there are mismatches in the voice or in the expressed behaviors by the robot, relative to the target speaker (e.g., in a tele-presence application where the target speaker' voice is reproduced by the robot), the impression of the individuality (or likeness) of the speaker is thought to be degraded.

So far, we have focused on generating natural human-like motions in the robot. However, no attention has been given to clarify how the robot's appearance or the robot's behaviors can influence the impression of the speaker individuality in android robots. We consider that the appearance, the voice and the behaviors by the robot are important factors for expressing the speaker individuality.

In this study, we investigated the effects of different modalities (including the voice, the robot's appearance and the robot's behaviors) for the impression of the speaker individuality. Two female-type androids are used for comparison, one is the android ERICA, while the other is the geminoid called totto, which is a copy of a famous Japanese TV talk show host. For the robot's behaviors, we focused on the effects of hand gestures, since we have observed in previous studies that the hand gesture types differ for different speakers [6].

The rest of paper is structured as following: Section II introduces prior work related to speaker individuality and gesture generation. Section III explains the content of the experiments conducted in this study and the generation method applied in the motions used in the experiments. Section IV provides a description of the experimental results. Section V discusses the outcome of the study, and last but not least, Section VI concludes the paper.

II. BACKGROUND AND RELATED WORK

Firstly regarding gesture generation, there have been many studies that have attempted to generate appropriate gestures associated with speech. [7] proposed Behavior Expression Animation Toolkit (BEAT), which generates text information as input information and generates rules based on the information obtained from the research results on people's traditional behavior. McNeill[8] proposes a method to classify gestures into functions such as iconic and metaphoric. [9] tried to generate motions for CG agents by analyzing superordinate concepts of words using WordNet[10], focusing on metaphorical gestures. As a generation method using machine learning, there has been proposed a model that performs learning using a neural network using LSTM, and generates motion data of a suitable gesture from speech features of a spoken voice by MFCC[11]. Also, prosody information consisting of energy and pitch of speech is input, and learning and generation are performed with a recursive neural network using a sequence conversion model[12]. There are more and more studies on gesture generation, but there are very few focusing on individuality-related features.

Regarding personality (which are somehow related to individuality), several studies have been conducted to express changes in personality of virtual agents by changing factors such as the speed, frequency, and size of gestures using Big Five personality as an indicator[13], [14], [15]. [16] gets the favored result by making her personality adaptive in the dialogue. In addition, [17] made a robot that imitates the motion of a certain human, and tries to recognize the individual's personality. [18] determines the contents of the utterance, gaze, and the gesture, to express the personality of the robot, based on rules, so that the displayed character is similar to the character of the subject. These studies focus on the so-called personality aspect (Big Five indicator) and do not explicitly mention about individuality.

[19] accounts for individuality features in iconic gesture generation. They report that gesture generation models trained by individual data produces motions with higher likeability, competence and humanlikeness scores, in comparison to models trained by combining data of multiple individuals. However, the individuality itself (i.e. the likeness) was not evaluated.

The research field of human identification and biometrics has studied a method to identify an individual from motion information. Especially, as the survey in [20], an identification from gait patterns have been actively studied in computer vision. These studies successfully revealed which gait motion features contribute to the identification of an individual, but they did not investigate the individuality in interaction context. This paper tackles to know the effect of robot's motion in individuality expression during communication with the robot.

Some studies have tackled the individuality using a teleoperated android robot that resembles a living individual. In [21], it is investigated whether the personal presence of the operator who is the original of the android can be represented through teleoperation. But they did not focus on which modality affects the personal presence. In [22], a "Doppel teleoperation system" is designed to isolate several physical traits (such as voice, face, and motion of the operator) from conversation to investigate whether people identify their acquaintances without physical traits during conversation. Their result suggested that even only conversational content is useful to identify the acquaintance. But they did not study how the motion affects the identification of an individual.

In this study, we investigate how the appearance, the voice and the gesture modality affect the impression of individuality.

III. INDIVIDUAL-SPECIFIC GESTURE GENERATION

In this section we will explain how the individual-specific motions were prepared, in order to conduct the experiment to evaluate the speaker's individuality. We consider that it would be possible to evaluate individuality by handling people that everyone knows when conducting the experiment. For that purpose, we decided to use an android robot called totto (right panel in Fig. 2), which is a copy of Tetsuko Kuroyanagi (hereinafter, TK), a famous Japanese TV talk show host (left panel in Fig. 2).

Fig. 2. Famous TV talk show host (left) and its android copy totto (right). ©2017 totto production committee

For comparison, we use the android ERICA (shown in Fig. 1) which was created by modeling a non-existent human, and therefore is not associated with a specific person though her appearance. The voice of a female speaker in our multimodal dialogue database is used for comparison. This female speaker is a talkative research assistant in her 30s, and will be called "speaker A" hereinafter. Unlike ERICA, the appearance of totto leads to a strong association to the speaker TK (Tetsuko Kuroyanagi).

A. *Gesture analysis*

In this subsection, we describe the gesture analysis results for the speaker TK, which are used for modeling TK-specific gesture generation. Recordings of a TV program on which she interviews guests are used as data for analysis and motion generation. Videos of four programs are used, each having about 30 minutes in length. The analysis was conducted by transcribing the text for each video, and segmenting and labeling the gesture events. The distributions of the gestures classified according to the categories proposed by McNeill, for each video, are shown in Fig. 3.

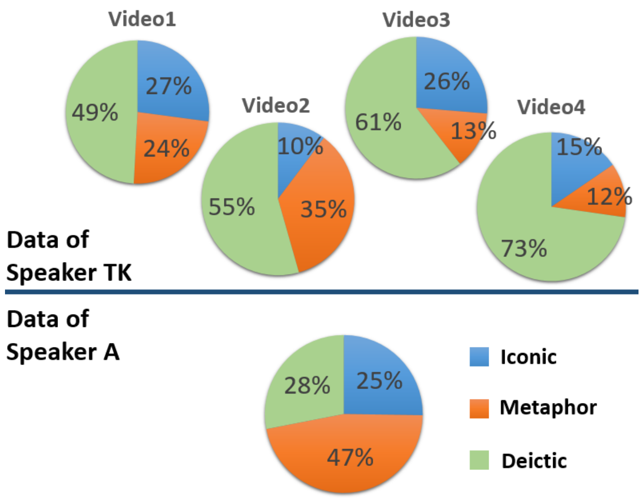

Fig. 3. Distributions of gesture categories for speaker TK and speaker A.

In this study, we focused on the following three gesture categories and calculate the occurrence rate of each. The beat gestures are omitted from this analysis since they usually co-occur with these gestures.

- Iconic: A gesture that represents a specific object or a specific shape.
- Metaphor: A gesture that represents an abstract concept.
- Deictic/Pointing: A gesture to indicate/point an opponent or an object

From these distributions, it was found that she had a high probability of generating pointing gestures compared to the distributions of other speakers reported in a previous study [6]. Therefore, it is considered that the movement itself also may reflect individuality. In addition, detailed analysis results on the characteristics of speaker TK's hand movement types are shown in the Table I. The number of times that both hands were used as gesture movements is Video 1: 7 (times), Video 2: 11 (times), Video 3: 6 (times), Video 4: 14 (times), Total: 38 (times).

From the analysis results of the motion types, it can be understood that there are more cases in which the gesture is performed with one hand than using both hands as a feature of speaker TK. This is considered to be because we usually use one hand rather than two hands when making movements that point to something such as a deictic gesture. Also, focusing on one-handed gestures, there are clearly more

TABLE I

DISTRIBUTIONS OF THE GESTURE MOVEMENT TYPES.

		a	b	c	d	e	f	g	h
Right Hand	Video1	7	12	8	6	8	2	0	6
	Video2	14	9	15	8	7	2	1	0
	Video3	0	3	4	1	2	1	12	2
	Video4	13	14	4	16	2	26	2	21
	Total	34	38	31	31	19	31	15	2
Left Hand	Video1	6	2	0	0	0	2	1	2
	Video2	3	2	0	0	1	4	0	0
	Video3	2	0	1	1	1	6	3	0
	Video4	5	0	0	0	0	11	0	1
	Total	16	4	1	1	2	23	4	3

The letters in the top of the table represents the type of movement, as follows: a: close finger-vertical palm to the ground, b: close finger-parallel palm to the ground, c: close finger-other, d: open finger-parallel palm to the ground-small vibrations, e: open finger-other, f: point with one index finger, g: point to self, h: other.

right-handed gestures. This is because of the arrangement of the guest seats in the TV program, where the guests are on the right side of speaker TK, and the right hand is increased because she frequently points to the guest to yield the turn. She also often performs instruction gestures in which the palm is parallel or perpendicular to the ground. By automatically generating these movements from the spoken text, it is expected that one could fully express the speaker TK.

B. *Gesture Generation*

As a method of automatically generating hand gestures, we use the method that has been developed in our previous study [6]. Fig. 4 shows the flow of the method. The method is based on text and prosodic information. Text information is used to generate gesture events (excluding beat gestures) by a Bayesian network associating word concepts and motion cluster data through gesture functions (such as iconic, metaphoric, and deictic). Prosodic information is used to generate beats at prosodic focus, and is superimposed with the text-based gestures. For details, refer to [6].

For training the models for the speaker TK, speech and text are extracted from the video data of the TV programs. However, the motion clusters are represented by a sequence of 3D points of the joints of the arms, and this information can not be straightly obtained from the 2D video data. We have tried to apply 3D skeleton models from 2D images, but the estimation accuracies were not sufficient. Therefore, in this study, the characteristic motions obtained from the analysis results of speaker TK data were picked up, and the Gesture Cluster in Fig. 4 were implemented by hand.

The gesture motion for speaker A was generated in the same way, by using the gesture cluster trained from the data of that speaker in [6]. It is worth to clarify that the only difference between the models of speakers A and TK are the way motion clusters were created, so that the generation processes are automatic for both speakers.

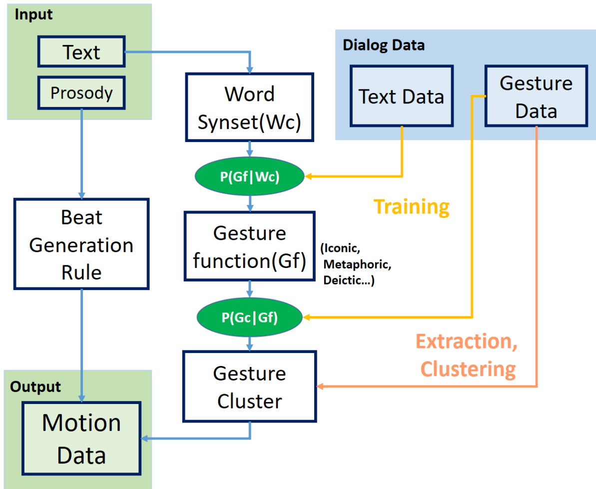

Fig. 4. Structure of Gesture Generation.

IV. EXPERIMENT METHOD AND RESULT

A. Experiment Setup

In this study, we conducted experiments to find out how much the different factors of appearance and hand gestures affect the impression of the individuality of a specific speaker. Therefore, we focused on appearance, hand motion, and speech, as elements expressing interaction with humans. For each of these three factors, we prepared data for two specific people and mixed the factors to evaluate which factors are related to the expression of speaker individuality. In short, in order to express the individuality of a specific speaker, we conducted experiments to investigate the effect of the most influential factor and the loss of consistency. The motion and voice / utterance contents are respectively generated from the video data of speaker TK and the recorded data of speaker A using the method described in Section III. For the appearance factor, in this experiment, two robots with different appearances (ERICA and totto) are used, as introduced in Section III.

TABLE II
GESTURE MOVEMENT ANALYSIS.

Factor	Condition
Robot(appearance)	ERICA or totto
Voice	speaker A or TK
Motion	speaker A or TK

Since a combination of these three factors are handled, experiments were performed under a total of eight conditions (2 robots x 2 voices x 2 motion types), as described in the Table II.

Segments of 1 to 2 minutes where the target speaker is predominantly talking were extracted from the data of both speaker A and speaker TK. Video clips were recorded for each of the eight conditions. For the overall motion generation, the previously proposed methods for lip motion [1], laughter motion [5], surprise motion [4], and the hand motion described in the previous section are employed. The experiment participants are asked to watch each video and answer to the following questionnaire. The order of the videos was randomized for the motion factor.

- Q1: Did the robot behave naturally (like a human)? (7 point scale: Unnatural(1) - Natural(7))
- Q2: Was the gesture natural? (7 point scale: Unnatural(1) - Natural(7))
- Q3: Was the amount of hand gesture appropriate? (7 point scale: Few(1) - Many(7))
- Q4: Did the robot behave like speaker TK? (7 point scale: Not like speaker TK(1) - Like speaker TK(7))
- Q5: Did the robot behave like speaker A? (7 point scale: Not like speaker A(1) - Like speaker A(7))

Seventeen subjects (7 male and 10 female, aged from 20s to 40s) participated in the evaluation experiments. All subjects know both the speaker A and speaker TK, so that they have an image of the personality and behavior of these two speakers.

B. Experiment Result

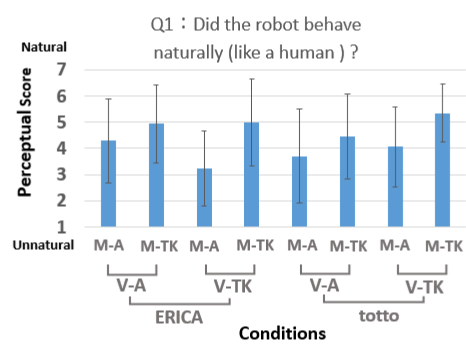

Fig. 5. Subjective naturalness of robot behavior.

Fig. 5 shows the average subjective scores of naturalness of robot behavior(Q1). The conditions in the horizontal axis are arranged by the robot appearance (ERICA or totto), the voice (V) and the motion type (M) for speakers A and TK. 3-way ANOVA was conducted on the influence of robot appearance, voice and motion type, on the motion naturalness (Q1). The analysis showed a main effect for motion type in the three-way interaction effect ($F(1,16) = 15.611$, $p = 0.001$, $\eta^2 = 0.494$). This means that the motions for the voice of speaker TK were significantly more natural than the motions for voice A. Overall, the speaker A gestures tended to be perceived as unnatural. The reasons for this will be discussed in the next section.

The results for Q2 regarding gesture naturalness are omitted, since similar trends with Q1 were observed. The results for Q3 regarding gesture amount are also omitted, since no significant differences were found among the factors.

Fig. 6 and 7 show the average subjective scores of likeness of speaker TK (Q4), likeness of speaker A (Q5). As shown in Fig. 6, the highest average score for likeness of speaker

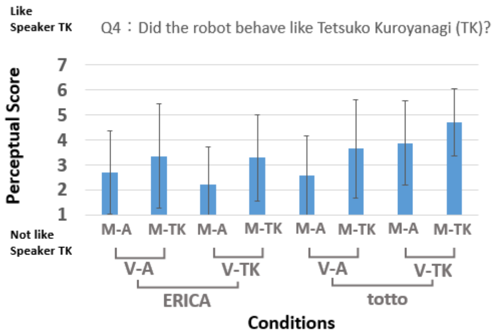

Fig. 6. Subjective likeness of speaker TK.

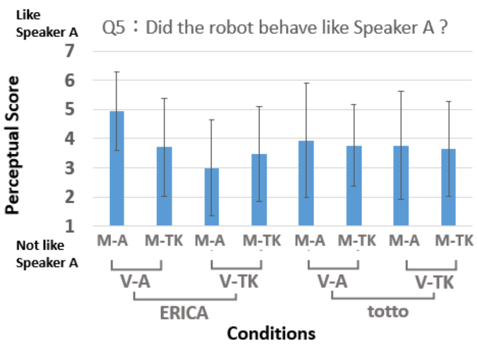

Fig. 7. Subjective likeness of speaker A.

TK was attributed for the condition where all factors (appearance, voice and motion) matched. The results on three-way ANOVA for Q4 (likeness of speaker TK) indicated main effects at 0.05 significance level for external appearance, and main effects at 0.01 significance level for motion type. In the appearance factor, the likeness of speaker TK is significantly higher in totto than in ERICA: $F(1,16) = 7.756$, $p = 0.013$, $\eta^2 = 0.326$. In the motion factor, the likeness of speaker TK is significantly higher for the motion of speaker TK (M-TK) than that of speaker A (M-A): $F(1,16) = 26.112$, $p = 0.0001$, $\eta^2 = 0.62$. The effects of voice and movement appeared to be similar, and increasing the mismatch resulted in a decrease in likeness of speaker TK.

The results in Fig. 7 show that the highest average score for likeness of speaker A was attributed for the condition where the voice, and the motion matched, for the android ERICA. For Q5 (likeness of speaker A), significant effect was found in the three-way interaction effect ($F(1,16) = 6.368$, $p = 0.023$, $\eta^2 = 0.285$). The main effects and two-way interaction effects were statistically not significant at a 0.05 significance level. The results of multiple comparisons with significant differences are as follows. When the appearance is ERICA and the motion type is of speaker A (M-A), the likeness of speaker A is higher for the voice of speaker A (V-A) than that of speaker TK (V-TK) ($p = 0.006$). When the appearance is ERICA and the voice is of speaker A (V-A), the likeness of speaker A is higher for the motion of speaker A (M-A) than that of speaker TK (M-TK) ($p = 0.022$).

From these results, it can be inferred that for ERICA, the

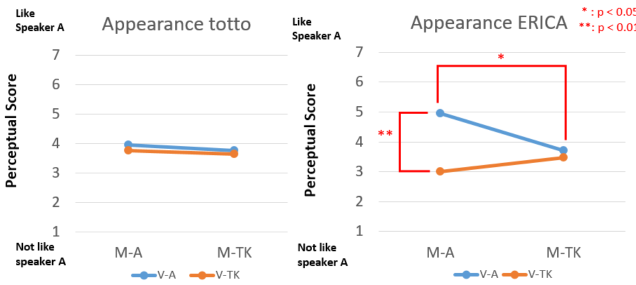

Fig. 8. Three-way interaction in likeness of speaker A (Q5).

likeness of speaker A decreases, if there is mismatch between voice and motion. On the other hand, no changes were found for totto. This is thought to be due to the strong appearance of totto in association to speaker TK. Overall, the results above indicate that both robot's appearance and motion types influence the impression of the speaker individuality.

V. DISCUSSION AND FUTURE WORK

In this experiment, it was found that the gesture generated by the data of speaker A was perceived as being unnatural compared to the motion of speaker TK. An explanation for this is that in the case of speaker TK, Gesture Clusters were manually generated, whereas in the case of speaker A, they were automatically generated from the sensor data, which included some noisy data. Motions created manually are still more natural than motions derived from sensing data. Another reason for lower naturalness scores for the speaker A is the lack of control of head and torso degrees of freedom. In previous evaluation, we have observed that the addition of the speaker's head and torso motions significantly increases the perceived naturalness. We have not included them in the present study, since this information was not available for the speaker TK's data. However, this also may be another factor influencing the impression of individuality. This is left for future investigation.

Significant differences were observed for external appearance and motion types in Q4, but not in Q5. A possible reason is that one of the robots is a copy of a person. The different trends observed in Q4 and Q5 suggest that the combination of modalities for the expression of individuality may differ for different people.

Moreover, as factors of expression of individuality dealt with in this study, the experiments were performed by taking appearance, voice and motion into account. However, it is thought that these factors may be further classified or the influence of other factors may be involved. For example, in terms of speech, the original voices of the speaker were used in this study, but the contents of speech also affect individuality. In fact, in [18], the experiment of expressing individuality is carried out in consideration of the amount of speech and the positive / negative of words. That is, even if the voice is the same, if the spoken contents are different, the impression of individuality might change. There are linguistic expressions that are characteristic of speaker A and of speaker TK, so it would be interesting to analyze those factors in more detail.

In addition, with regard to motion, experiments were performed by considering only the gesture accompanying the speech. However, human motion is more complicated, and it is common to perform chewing movements such as scratching the head or touching the nose. In addition to the movement, the attitude during the dialogue (like setting up arms, leaning on the armrest, etc.) also leads to the expression of the humanlikeness, and the frequency of the change in the attitude also affects the expression of individuality.

Therefore, it can be said that the experiment conducted in this study reflects only part of the human action factors. In the future, it will be necessary to implement such factors regarding habit and attitude that are not highly dependent on speech, and to investigate the expression of individuality. Right now, we are generating gestures on a word basis, but if we use a generation method that takes into account situations, emotions, etc., the humanlikeness and individuality perception must increase further.

VI. CONCLUSIONS

In this study, we investigated the influence of three factors, appearance, voice and hand motion, on the impression of the speaker individuality in android robots. Combination of two voices (speaker A and speaker TK), two android robots (ERICA and totto, which is a copy of speaker TK), and two hand motion types (generated by hand gesture models trained for each of the speakers A and TK) were evaluated to check how each modality affects the impression of the speaker individuality.

Evaluation results indicated that all these three factors affect the impression of speaker individuality. The highest scores for the speaker likeness were obtained when all modalities are matched, while the impressions are decreased, when modality mismatches occur. This indicates that the hand motion modality is an important factor for expressing the speaker individuality, besides the appearance and the voice.

Further, different trends were found depending on whether or not the android is copy of an existent person. The android totto, which has a strong image of speaker TK, is not so appropriate for expressing individuality features of other people, while the android ERICA, which is not a copy of an existent person, can express better the speaker individuality if the modalities match with the image of that individual.

In future, we intend to check the effects of other modalities such as head, torso and eye gazing features on individuality.

ACKNOWLEDGMENT

We thank Dentsu Tec Inc. for lending the android robot totto, and the TV Asahi Corporation for providing the recording data of Tetsuko Kuroyanagi TV program. We also thank Yumiko Baba, Taeko Murase, and Megumi Taniguchi for contributions in the data annotation and experiment conduction.

REFERENCES

[1] C. T. Ishi, C. Liu, H. Ishiguro, and N. Hagita, “Evaluation of formant-based lip motion generation in tele-operated humanoid robots,” in *IROS*, 2012, pp. 2377–2382.

[2] C. Ishi, C. Liu, H. Ishiguro, and N. Hagita, “Head motion during dialogue speech and nod timing control in humanoid robots,” in *Proc. ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*, 2010, pp. 293–300.

[3] C. Liu, C. T. Ishi, H. Ishiguro, and N. Hagita, “Generation of nodding, head tilting and gazing for human –robot speech interaction,” *Int. J. of Humanoid Robotics*, vol. 10, no. 01, p. 1350009, 2013.

[4] C. T. Ishi, T. Minato, and H. Ishiguro, “Motion analysis in vocalized surprise expressions and motion generation in android robots,” *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1748–1754, 2017.

[5] C. Ishi, T. Minato, and H. Ishiguro, “Analysis and generation of laughter motions, and evaluation in an android robot,” *APSIPA Transactions on Signal and Information Processing*, vol. 8, 2019.

[6] C. T. Ishi, D. Machiyashiki, R. Mikata, and H. Ishiguro, “A speech driven hand gesture generation method and evaluation in android robots,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3757–3764, 2018.

[7] J. Cassell, H. H. Vilhjálmsson, and T. Bickmore, “Beat: the behavior expression animation toolkit,” in *Life-Like Characters*. Springer, 2004, pp. 163–185.

[8] D. McNeill, *Hand and mind: What gestures reveal about thought*. University of Chicago press, 1992.

[9] Y. Kadono, Y. Takase, and Y. Nakano, “Analyzing metaphoric gestures towards automatic gesture generation,” *The 29th Annual Conf of the Japanese Society for Artificial Intelligence*, vol. 29, pp. 1–3, 2015.

[10] F. Bond, T. Baldwin, R. Fothergill, and K. Uchimoto, “Japanese semcor: A sense-tagged corpus of japanese,” *Proceedings of the 6th Global Wordnet Conference(GWC 2012)*, pp. 56–63, 2012.

[11] D. Hasegawa, N. Kaneko, S. Shirakawa, H. Sakuta, and K. Sumi, “Evaluation of speech-to-gesture generation using bi-directional lstm network,” in *Proc. Int. Conf. on Intelligent Virtual Agents*, 2018, pp. 79–86.

[12] A. Shimazu, C. Hieida, T. Nagai, T. Nakamura, Y. Takeda, T. Hara, O. Nakagawa, and T. Maeda, “Generation of presentation motion using sequence to sequence model,” *The 32nd Annual Conference of the Japanese Society for Artificial Intelligence*, vol. 32, 2018.

[13] K. Liu, J. Tolins, J. E. Fox Tree, M. Neff, and M. A. Walker, “Two techniques for assessing virtual agent personality,” *IEEE Trans. Affect. Comput.*, vol. 7, no. 1, pp. 94–105, Jan. 2016.

[14] M. Neff, Y. Wang, R. Abbott, and M. Walker, “Evaluating the effect of gesture and language on personality perception in conversational agents,” in *Int. Conf. on Intelligent Virtual Agents*, 2010, pp. 222–235.

[15] B. Hartmann, M. Mancini, and C. Pelachaud, “Implementing expressive gesture synthesis for embodied conversational agents,” in *International Gesture Workshop*. Springer, 2005, pp. 188–199.

[16] C. Hu, M. A. Walker, M. Neff, and J. E. F. Tree, “Storytelling agents with personality and adaptivity,” in *International Conference on Intelligent Virtual Agents*. Springer, 2015, pp. 181–193.

[17] P. Bremner, O. Celiktutan, and H. Gunes, “Personality perception of robot avatar tele-operators,” in *HRI*, 2016, pp. 141–148.

[18] A. Aly and A. Tapus, “A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction,” in *Proc. HRI*, 2013, pp. 325–332.

[19] K. Bergmann, S. Kopp, and F. Eyssel, “Individualized gesturing outperforms average gesturing–evaluating gesture production in virtual humans,” in *Int. Conf. on Intelligent Virtual Agents*, 2010, pp. 104–117.

[20] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, “The humanid gait challenge problem: data sets, performance, and analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 162–177, 2005.

[21] S. Nishio, H. Ishiguro, and N. Hagita, “Can a teleoperated android represent personal presence? a case study with children,” *Psycologia*, vol. 50, no. 4, pp. 330–342, 2007.

[22] H. Sumioka, S. Nishio, E. Okamoto, and H. Ishiguro, “Doppel teleoperation system: Isolation of physical traits and intelligence for personality study,” in *Proc. of Annual meeting of the Cognitive Science Society*, 2012, pp. 2375–2380.