Effect of a Hearer’s Politeness on Multimodal Behaviors in Speech

Noriko Suzuki*, Haruka Shoda**, Rina Matsubayashi***, and Mamiko Sakata***

Abstract: This paper investigates the effect of a hearer’s attitude toward a speaker’s multimodal behaviors. Twenty-one university students, nine males and twelve females, participated in this experiment as speakers. Two females, aged 22 years old, performed in the role of hearers. The experiment was carried out under two conditions regarding the hearer’s attitude: polite and impolite conditions. The polite hearer nodded, produced back-channel response, and gazed at the speaker, while the impolite hearer did these actions to a lesser extent. Under each condition, the speakers talked about their own experiences to the two different hearers through a within-participants design. The speech, gaze, and body motion of speakers were recorded with a video camera. In addition, the communication skills of speakers, including expressivity, sensitivity, regulation, assertiveness, and responsiveness skills, were measured after the experiment. The speakers’ silent pauses, filled pauses, gaze aversion, and representational gestures were compared between the hearers’ two politeness conditions. The relationship between the multimodal behaviors and the speakers’ communication skills was also analyzed. The results suggest that the speakers’ silent pauses, filled pauses, and gaze aversion decreased in the polite hearer condition, while representational gestures increased. From the communication-skills test, the speakers with higher expressivity, sensitivity, and regulation skills but lower assertiveness and responsiveness skills show longer silent pauses under the impolite hearer condition. Moreover, we discuss how the implications of our findings can enhance the relationship between social robots and people.

Key Words: politeness, multimodal behaviors, communication skills, narrative interaction.

1. Introduction

This paper focuses on the change in a conversation participant’s multimodal behaviors when the other participant changes her/his attitude toward the partner in face-to-face interaction between two people.

In our everyday life, people change their behaviors in response to what they see their communication partners doing. People’s adaptation to what a communication partner does has been observed to lead to changes in their emerging behavior [1]. This paper investigates the relationship between a communication partner’s attitude and people’s multimodal behaviors, including silent and filled pauses, gaze, and gestures. These multimodal behaviors of people can work as behavioral indexes of the relationship between people and their communication partners, such as speakers and hearers, when these behaviors change in response to a different attitude of the partners.

Silent and filled pauses are typical examples of paralinguistic cues for hesitation markers, or they may indicate disfluency in spontaneous speech. Speakers can use these pauses as announcements to hearers that they are searching for a word or are deciding what to say next (e.g., [2], [3]). Furthermore, these pauses can be considered potential social markers of speakers relative to the hearers’ attitude.

Gaze is continuously expressed in interaction. Gaze has been said to have a number of different functions, e.g., information seeking, signaling interpersonal attitudes, controlling the synchronizing of speech, expressing intimacy, and displaying inhibition to avoid intimacy or distraction [4]. Many conventional studies have focused on eye contact as an indicator of the hearer’s or a counselor’s interest in the speaker or a client [5]–[7]. On the other hand, gaze aversion appears during the speaker’s pauses of hesitation when it is necessary to concentrate on the beginning of the speech [8].

Gesture, one of the body motions, is also a continuous expression throughout interaction. Gesture is classified into several types, such as emblem and spontaneous gesture (e.g., [9], [10]). Representational gesture is one of the spontaneous gestures, and it illustrates the internal image of the speaker, including the location and the event [9].

This paper examines the effects of the hearer’s attitude, either polite or impolite, on the multimodal behaviors through the task of experience telling. In this task, people talk about their own experience, either happy or troubled, to a “hearer” actress. The polite hearer expresses herself simply by nodding, producing back-channel responses, and looking into the speaker’s face in an appropriate manner throughout the task. On the other hand, the impolite hearer only sometimes expresses herself by nodding, only sometimes produces a back-channel response, and averts her gaze. Neither the polite nor the impolite hearer returns any linguistic response to the speaker.

We also focus on the relationship between multimodal behaviors and communication skills. There is much related work in this area, e.g., silent or filled pauses [11], [12], gaze aversion [13], representational gestures [14], head orientation, nods, leg movement, and leg extension [15]. In particular, the af-
fective communication test is a typical example of assessing nonverbal emotional expressiveness [16]. Moreover, some conventional studies in human-robot interaction have attempted to predict the level of engagement in profiling people's personality traits, as either extroversion or introversion [17]–[19].

These conventional studies investigated the correlative relationship between nonverbal behaviors and communication skills. This paper examines the causal relationship of communication skills, i.e., as measured by ENDCOREs (ENcode, Decode, CONtrol and REGulate) [20], and multimodal behaviors using path analysis. ENDCOREs has been constructed to measure the social competence of multimodal behaviors of people in face-to-face communication independent of culture or society. It integrates 11 existing approaches to measuring communication skills, e.g., PEA (Perceived Encoding Abilities)/PDA (Perceived Decoding Abilities) [21], ACT (the Affective Communication Test) [16], SSI (Social Skills Inventory) [22], SSI for Japanese, ICQ (Interpersonal Competence Questionnaire) [23], ICQ for Japanese [24], KiSS (Kikuchi's Social skill Scale: 18 items) [25], JICS (Japanese Interpersonal Communication Competence) [26], RSMS (Revised Self-Monitoring Scale) [27]. The ENDCOREs model was constructed using 233 Japanese university students [20] and verified using another 2,184 Japanese university students [28]. Consequently, the authors used only ENDCOREs as a means of measuring the communication skills of participants. ENDCOREs covers the six communication skills of self-control, expressivity, sensitivity, assertiveness, responsiveness, and regulation [29], [30].

From studying the above related work [5]–[7], [11]–[14], we predicted that both silent and filled pauses as well as gaze aversion will increase, while the amount of speech and representational gestures will decrease when the speaker talks to an impolite hearer, and vice versa in the case of a polite hearer.

In addition, regarding the relationship between communication skills and multimodal behaviors, we predicted that more favorable traits will promote representational gestures, while less favorable traits will promote silent/filled pauses or gaze aversion.

2. Method

2.1 Experimental Design

The experiment was conducted using a within-subjects design with the hearer's politeness as a factor. In the experiment, a participant (a speaker) related his/her own experiences to a hearer actress, who reacted to the speaker in either a polite or an impolite manner. In the next condition, the same speaker talked to another hearer, who reacted to him/her in the manner opposite to the first hearer. We assessed the speaker’s multimodal behaviors (for example, the amounts of speech, silent pause, filled pause, gazing, gesture) in the polite versus impolite conditions, and then we explored how such behavioral manipulations were determined by the speaker’s communication skills.

2.2 Participants

As subjects, 21 undergraduates (9 men, 12 women) aged 18 to 21 (\(M = 19.81, SD = 1.25\)) participated as speakers in the experiment. Written informed consent was obtained from all participants.
satisfaction with their own speech on a questionnaire immediately after each condition. The questionnaire was created based on a previous study [31]. Participants rated the extent to which the speech was well-coordinated, cooperative, awkward, involving, intense, and friendly, using an 8-point Likert-type scale for 14 items (1: Not at all to 8: Extremely). The summed score (Cronbach’s α = .92) was used in the analysis after processing the reverse items.

After completing the self-evaluation of the speech, participants were also asked to self-evaluate their own communication skills using the ENDCOREs questionnaire [20]. ENDCOREs enabled us to measure the six components of communication skills—self-control, expressivity, sensitivity, assertiveness, responsiveness, and regulation—using a 7-point Likert scale for 24 items (1: Very poor to 7: Very Good). In other words, ENDCOREs is capable of measuring the social competence of people [29], [30]. For each of these measures, a higher score indicates better communication skills.

These communication skills do not seem to be easily influenced by short-term interaction, as with other personality traits including the Big Five. Personality traits such as communication skills are stable for people unless they receive an explicit type of skills training.

2.4 Behavior Coding

For the three-minute recording in each condition per participant, we extracted multimodal behaviors using the annotation software ELAN (EUDICO Linguistic Annotator [32]). In the following analyses, we used each ratio for the amount of speech, silent pause, filled pause, gaze toward the hearer, gaze aversion, and representational gestures, since these parameters are assumed to be affected by the total duration of speech.

2.4.1 Speech (in second)

Total amount of speech units (seconds) per three-minute interaction. A speech unit is the duration of a speaker’s speech bounded by pauses. The speech unit includes back-channel responses, e.g., hai (yeah), sou sou (that’s right).

2.4.2 Silent pause (in second)

Total amount of silent pause units (seconds) per three-minute interaction. A silent pause unit is the duration of a speaker’s silence bounded by speech.

2.4.3 Filled pause (in second)

Ratio of filled pause units per three-minute interaction. A filled pause unit is a hesitation marker or disfluency in spontaneous speech indicated with a sound, e.g., ehto (er) or anoo (uh). Filled pause is used when speakers are searching for a word or are deciding what to say next (e.g., [2], [3]).

2.4.4 Gaze toward the hearer (in second)

Ratio of duration of speaker’s gaze to the hearer per three-minute interaction. The directions of the speaker’s gaze are estimated from the direction of the eye’s dark region of the pupil and iris. The gaze units are divided into two categories: the hearer and other directions.

2.4.5 Gaze aversion (in second)

Ratio of duration of speaker’s gaze aversion from the hearer per three-minute interaction. The gaze unit of the speaker is directed in another direction than toward the hearer.

2.4.6 Representational gesture (in second)

Ratio of speaker’s representational gesture units per three-minute interaction. Representational gestures express seman-
value was greater for the speaker with more assertiveness and responsiveness but less expressivity, sensitivity, and regulation. This indicates that people who have such communication skills tended to pause silently more in the polite than in the impolite condition. Ratios of filled pause and representational gesture to the amount of speech yielded similar models: The speakers with more assertiveness tended to show more filled pause and gestures in the polite than in the impolite condition.

Multiple regression analysis is a useful way to understand how each of the independent variables (i.e., communication skills) affects one dependent variable (the speaker’s behavior), whereas it cannot be applied to the data in which we assumed correlations among variables (as in the case of “multicollinearity”) or multiple dependent variables such as our data. To solve this problem, and to visualize the causal relationship, we conducted path analysis (Fig. 2) based on the variables selected in the multiple regression analyses (Table 2). A path analysis is a statistical method used to describe the directed dependencies among a set of observed variables, which enables us to reveal the interrelationships among independent and dependent variables [36]. Here, we excluded one variable, “amount of speech,” because this value highly correlated with the “amount of silent pause” variable ($r = -0.93$, $p < .001$). We tested the graphical models by using the generalized least squares method, appropriate for small samples such as ours ($N = 21$) in which the assumption of multivariate normal distribution cannot be hypothesized [37]. By removing insignificant paths, the obtained model (Fig. 2) was confirmed to fit the current data by the criteria of goodness-of-fit indices: $\chi^2(8, N = 21) = 3.02$, $p = .93$, $GFI = .96$, $AGFI = .83$, $CFI > .99$, $RMSEA < .001$. $GFI$, $AGFI$, $CFI$, and $RMSEA$ stand for goodness-of-fit index, adjusted goodness-of-fit index, comparative fit index, and root mean square error of approximation, respectively. The final model (Fig. 2) shows that the amount of silent pause from the polite to impolite condition was determined by less expressivity, sensitivity, and regulation skills as well as by more assertiveness and responsiveness skills of the speakers. From another viewpoint, the speakers with skills of expressivity, sensitivity, and regulation tended to pause silently more in their speech when the hearer was impolite; in contrast, the speakers with assertiveness and responsiveness skills were not likely to pause silently more in the impolite condition. The assertive speakers were not likely to show more filled pause in the impolite rather than in the polite condition, while they were likely to show more representational gestures from the polite to the impolite condition.

In summary, we showed that the multimodal behaviors of university students were fully affected by the hearer’s attitude of politeness. An impolite reaction by the hearer made the speakers pause silently more, along with using more filled pause and gaze aversion but fewer representational gestures. The present study shows that one of the factors of such differences was the speaker’s communication skills. The tendency of “paused silently more in the impolite condition” was likely to be shown by the variables removed by a stepwise procedure based on Akaike’s information criteria.

### Table 1
Mean (and standard deviation) for each variable in each condition. Results of paired $t$-tests are shown in right columns.

| Parameters                          | Polite Condition | Impolite Condition | $t(20)$ | $p$ | $d$ |
|-------------------------------------|------------------|--------------------|---------|----|-----|
| Amount of speech (s)                | 113.30 (22.55)   | 99.41 (24.22)      | 5.78    | <.001 | 0.59 |
| Amount of silent pause (s)          | 73.98 (22.52)    | 86.04 (23.18)      | 4.63    | <.001 | 0.53 |
| Ratio of filled pause               | 0.05 (0.02)      | 0.06 (0.04)        | 2.19    | .04  | .42  |
| Ratio of gaze toward hearer         | 0.73 (0.47)      | 0.58 (0.35)        | 2.27    | .03  | .35  |
| Ratio of gaze aversion              | 0.91 (0.37)      | 1.31 (0.57)        | 4.30    | <.001 | 0.84 |
| Ratio of representational gesture   | 0.38 (0.23)      | 0.26 (0.21)        | 3.29    | .004 | 0.50 |

$^p < .001$, $^* p < .01$, $^*^* p < .05$

### Table 2
Results of multiple regression analyses determining effects of speaker’s skills on communication behaviors. As communication behaviors, difference scores from polite to impolite conditions were used. CS1 to CS6 indicate constructs of EENDCOREs (self-control, expressivity, sensitivity, assertiveness, responsiveness, and regulation, respectively). “-” indicates that the variable was removed by a stepwise procedure based on Akaike’s information criteria.

| Dependent variable                  | CS1     | CS2     | CS3     | CS4     | CS5     | CS6     |
|-------------------------------------|---------|---------|---------|---------|---------|---------|
| Amount of speech                    | -.49    | .53     | -.24    | -.61    | .34     |
| Amount of silent pause              | -.52    | -.49    | .26†    | .52†    | -.40‡   |
| Ratio of filled pause               |         |         | .39†    |         |         |
| Ratio of gaze toward hearer         |         |         |         |         |         |
| Ratio of gaze aversion              | -.32    | -.29    |         |         |         |
| Ratio of representational gesture   |         |         |         | .39†    |         |

$^* p < .001$, $^*^* p < .01$, $^*^*^* p < .05$, $^* p < .10$

### Fig. 2
Path diagram explaining effects of speaker’s skills on communication behaviors. Dependent variables were computed by the difference from polite to impolite conditions. Solid and dashed lines indicate positive and negative coefficients, respectively. Boldness of lines reflects degree of significance. e1 to e3 indicate error variables.
by speakers with higher expressivity, sensitivity, and regulation skills but with lower assertiveness and responsiveness skills. The assertiveness skill suppressed the speaker’s use of filled pause in the impolite condition while promoting the use of representational gesture in the polite condition.

4. Discussion

4.1 Relationship between Hearer’s Politeness and Speaker’s Multimodal Behaviors

Both the amount of silent pauses and the ratio of filled pauses per speech increased in the impolite hearer condition more than in the polite hearer condition. From the conventional studies, silent and filled pauses have been considered a kind of disfluent behavior as a speaker-based stance (e.g., [2], [3]). It seems that speakers have used silent and filled pauses to announce their conversational status, i.e., their intention to start or continue to speak. On the other hand, from our results, silent and filled pauses seem to be affected by a change in the hearer’s attitude. In other words, silent and filled pauses might play the role of expressing a hearer-oriented stance.

Similarly, representational gestures also seem to be produced by a speaker’s internal motivation (e.g., [9], [10]). From our results, the ratio of representational gestures increased in the polite hearer condition more than in the impolite hearer condition. The change of a hearer’s attitude seems to affect the production of a speaker’s representational gestures. Representational gestures might also play the role of expressing hearer-oriented behavior.

On the other hand, gaze seems to play both roles, that is, expressing speaker-based and hearer-oriented stances. Both gaze and gaze aversion have been considered ways of expressing the social relationships among people, including the level of intimacy or the process of turn-taking (e.g., [4]–[8]). From our results, the ratio of gaze aversion increased in the impolite hearer condition more than in the polite hearer condition, although the ratio of gaze toward the hearer showed the opposite trend. Our results thus appear to strongly endorse the findings of conventional studies of gaze in conversation.

From these results, multimodal behaviors, silent and filled pauses, gaze aversion, and representational gestures of speakers might also play the role of expressing hearer-oriented stances.

4.2 Relationship between Speaker’s Communication Skills and Their Multimodal Behaviors

Both silent and filled pauses have been pointed out as reflecting a speaker’s anxiety (e.g., [11]), although no relation has been found between the Big Five personality traits and the use of filled pauses (e.g., [12]). On the other hand, we focus on the relationship between communication skills and multimodal behaviors. We selected ENDCOND to measure the communication skills of people, since it is specifically designed for the skills of interpersonal relations [20]. Through path analysis, we found a tendency of more frequent expression of silent pauses in the impolite condition by speakers with higher expressivity, sensitivity, and regulation skills but lower assertiveness and responsive skills. In addition, the lower assertiveness skill suppressed the speaker’s use of filled pauses in the impolite condition. From these results, the five social skills of speakers might influence the expression of the amount of silent pauses. Similarly, the ratio of filled pauses might be suppressed by the assertiveness skill of speakers. Silent and filled pauses of speakers might play the role of expressing a speaker-oriented stance.

Conventional studies of gaze aversion have pointed out that it is associated with unfavorable traits such as shyness, social anxiety, and negative social evaluation with less sincerity (e.g., [13]). However, we did not find any explicit influence of social skills on gaze aversion as a result of path analysis.

On the other hand, representational gestures have been pointed out as having a positive correlation with extroversion and neuroticism in the Big Five personality traits (e.g., [14]). Our path analysis found that the ratio of representational gestures might be impaired by the assertiveness skill of speakers. Representational gestures might also involve the role of expressing a speaker-oriented stance.

4.3 Relationship between Social Robots and People

Some conventional studies have tried to predict automatically the personality traits of people through tracking their behaviors in human-robot interaction, including voice prosody [17], motion, appearance, visual attention [19], distance, synchrony, and dominance [18]. These results would apply to social robots producing appropriate behaviors according to the personal traits of people.

In this paper, we focused on other multimodal behaviors: the amount of speech, silent and filled pauses, gaze aversion from hearers, and representational gestures. We obtained a path diagram from communication skills to multimodal behaviors. These behaviors can be detected in real time using audio and visual sensors as well as reference to the nonverbal behaviors of conventional studies in human-robot interaction [17]–[19]. Our results will be applied to automatically predicting the communication skills of people, such as the work done in Youyou et al. (2015) [38].

Our findings can help to improve the design of minimal relationality between people and social robots that have a function for controlling the degree of politeness while monitoring behaviors from human partners (e.g., [39], [40]). In other words, our results could be applied to a mechanism of social robots to express attitude of minimal politeness according to the level of communication skills of human partners for constructing and maintaining relationship with them.

5. Conclusion

The purpose of this study was to investigate the effect of a hearer’s attitude toward a speaker’s multimodal behaviors.

The experiment was carried out under two conditions of the hearer’s attitude: polite and impolite conditions. The polite hearer nodded, produced back-channel responses, and gazed at the speaker, while the impolite hearer made fewer such actions. The speakers talked about their own past experiences under each condition to two different hearers through a within-participants design.

The speech, gaze, and body motion of speakers were recorded with a video camera. A communication-skills test of speakers, including expressivity, sensitivity, regulation, assertiveness, and responsiveness skills, was conducted after the experiment.

The results show that the speakers’ silent pauses, filled pauses, and gaze aversion decrease in the polite hearer condi-
tion, while representational gestures increase. The results suggest that these multimodal behaviors play the role of expressing a hearer-oriented stance.

As for the relationship between multimodal behaviors and communication skills, the results from path analysis indicate that speakers with higher expressivity, sensitivity, and regulation skills but lower assertiveness and responsiveness skills show longer silent pauses in the impolite hearer condition. The assertiveness skill suppressed filled pauses by the speakers in the impolite hearer condition while it enhanced representational gestures. The results suggest that these multimodal behaviors also play the role of expressing a speaker-oriented stance.

As our future work, we will execute more detailed analysis, i.e., investigating the temporal structures among multimodal behaviors as done by Harrigan (1985) [7], in order to map these behaviors to personal traits [15] using the ENDCORE measurement of communication skills [20]. We will also confirm whether people produce different amounts of multimodal behaviors when they interact with a computer or a robot with polite/impolite reactions, as done by Reeves and Nass (1996) [41]. Furthermore, we will examine the level of acceptance by people, according to the level of communication skills of people, while changing the level of impoliteness of the computer or robot; such study will help us determine the minimal design of a robot’s polite attitude to construct and maintain relations with people. In other words, we should clarify the range of people’s acceptance of social robots’ attitudes including impoliteness through human-robot interaction.

Acknowledgments

The findings of this study are based on the third author’s senior thesis. We thank 21 participants and two actresses at Doshisha University for their participation in the experiment. This work was supported by JSPS KAKENHI Grant Numbers JP17K01653 and JP15K00219.

References

[1] C. Goodwin: Embodied hearers and speakers constructing talk and action in interaction, Cognitive Studies, Vol. 16, No. 1, pp. 51–64, 2009.
[2] H.H. Clark and J.E.F. Tree: Using uh and um in spontaneous speaking, Journal of Cognition, Vol. 84, No. 1, pp. 73–111, 2002.
[3] J. Yuan, X. Xu, W. Lai, and M. Liberman: Pauses and pause fillers in Mandarin monologue speech: The effects of sex and proficiency, Proc. Speech Prosody, pp. 1167–1170, 2016.
[4] M. Argyle, R. Ingham, F. Alkema, and M. McCallin: The different functions of gaze, Semiotica, Vol. 7, No. 1, pp. 19–32, 1973.
[5] O. Hargie, C. Saunders, and D. Dickson: Social Skills in Interpersonal Communication, 3rd Edition, Routledge, 1994.
[6] A. Ivey, C.J. Normington, C.D. Miller, and R.F. Haase: Microcounseling and attending behavior: An approach to practicum counselor training, Journal of Counseling Psychology, Vol. 15, No. 5 (Pt 2), pp. 1–12, 1968.
[7] J.A. Harrigan: Listeners’ body movements and speaking turns, Communication Research, Vol. 12, No. 2, pp. 233–250, 1985.
[8] A. Kendon: Some functions of gaze-direction in social interaction, Acta Psychologica, Vol. 26, pp. 22–63, 1967.
[9] D. McNeill: Hand and Mind: What Gestures Reveal about Thought, Univ. of Chicago Press, 1992.
[10] S. Kita: How representational gestures help speaking, McNeill, Ed., Language and Gesture: Window into Thought and Action, pp. 162–185, Cambridge Univ. Press, 2000.
[11] S.R. Rochester: The significance of pauses in spontaneous speech, Journal of Psycholinguistic Research, Vol. 2, No. 1, pp. 51–81, 1973.
[12] C.M. Laserna, Y.T. Seih, and J.W. Pennebaker: Um ... who like says you know: Filler word use as a function of age, gender, and personality, Journal of Language and Social Psychology, Vol. 33, No. 3, pp. 328–338, 2014.
[13] R.J. Larsen and T.K. Shackelford: Gaze avoidance: Personality and social judgments of people who avoid direct face-to-face contact, Personality and Individual Differences, Vol. 21, No. 6, pp. 907–917, 1996.
[14] A.B. Hostetter and A.L. Pothoff: Effects of personality and social situation on representational gesture production, Gesture, Vol. 12, No. 1, pp. 63–84, 2012.
[15] R. Gifford: Mapping nonverbal behavior on the interpersonal circle, Journal of Personality and Social Psychology, Vol. 61, No. 2, pp. 279–288, 1991.
[16] H.S. Friedman, L.M. Prince, R.E. Riggio, and M.R. DiMatteo: Understanding and assessing nonverbal expressiveness: The affective communication test, Journal of Personality and Social Psychology, Vol. 39, No. 2, pp. 333–351, 1980.
[17] G. Mohammad and A. Vinciarelli: Automatic personality perception: Prediction of trait attribution based on prosodic features, IEEE Trans. on Affective Computing, Vol. 3, No. 3, pp. 273–284, 2012.
[18] S.M. Anzalone, G. Varni, E. Zibetti, S. Ivaldi, and M. Chetouani: Automated prediction of extraversion during human-robot interaction, International Journal of Social Robotics, Vol. 9, No. 3, pp. 385–399, 2017.
[19] H. Salam, O. Celikutkutan, and I. Hupont: Fully automatic analysis of engagement and its relationship to personality in human-robot interactions, IEEE Access, Vol. 5, pp. 705–721, 2017.
[20] M. Fujimoto and I. Daibo: ENDCOREs: A hierarchical structure theory of communication skills. The Japanese Journal of Personality, Vol. 15, No. 3, pp. 347–361, 2007 (in Japanese).
[21] M. Zuckerman and D.T. Lannace: Individual differences in perceived encoding and decoding abilities, R. Rotenthal (Ed.), Skill in Nonverbal Communication: Individual Differences, pp. 171–203, Oelgeschlager, Gunn & Hain, 1979.
[22] R.E. Riggio: Assessment of basic social skills, Journal of Personality and Social Psychology, Vol. 51, pp. 649–660, 1986.
[23] D. Burmester, W. Furman, M.T. Wittenberg, and H.T. Reis: Five domains of interpersonal competence in peer relationships, Journal of Personality and Social Psychology, Vol. 55, pp. 991–1008, 1988.
[24] J. Kayano: An integrated approach to social skills research (I): Examination of consistency and validity of SSI, Human Science the Graduate Course of Kansai University, Vol. 31, pp. 1–16, 1988 (in Japanese).
[25] A. Kikuchi: Science of Caring, Kawashima-Shoten, 1988 (in Japanese).
[26] J. Takai and H. Ota: Assessing Japanese interpersonal communication competence, Japanese Journal of Experimental Social Psychology, Vol. 33, pp. 224–236, 1994.
[27] R.D. Lennox and R.N. Wolf: Revision of the self monitoring scale, Journal of Personality and Social Psychology, Vol. 46, pp. 1349–1364, 1984.
[28] M. Fujimoto: An empirical and conceptual examination of the ENDCORE model for practical work with communication skills, The Japanese Journal of Personality, Vol. 22, No. 2, pp. 156–167, 2013 (in Japanese).
[29] T. Anme, T. Watanabe, K. Tokutake, E. Tomisaki, Y. Mochizuki, E. Tanaka, B. Wu, M. Nanba, R. Shinohara, and Y. Sugisawa: A pilot study of social competence assessment using Interaction Rating Scale Advanced, ISRN Pediatrics, Vol. 2011, Article ID 272913, 2011.
[30] A. Komazawa and I. Ishimura: Strengthspotting and interpersonal relationships: Development of the Japanese version of the strengthspotting scale, *GSTF Journal of Psychology*, Vol. 2, No. 2, pp. 29–36, 2016.

[31] M. Kimura, M. Yogo, and I. Daibo: Expressivity halo effect in the conversation about emotional episodes, *Japanese Journal of Research on Emotions*, Vol. 12, No. 1, pp. 12–23, 2005 (in Japanese).

[32] H. Lausberg and H. Sloetjes: Coding gestural behavior with the NEUROGES-ELAN system, *Behavior Research Methods*, Vol. 41, No. 3, pp. 841–849, 2009.

[33] W.N. Venables and B.D. Ripley: *Modern Applied Statistics with S*, 4th Edition, Springer, 2002.

[34] Y. Rosseel: lavaan: An R package for structural equation modeling, *Journal of Statistical Software*, Vol. 48, No. 2, pp. 1–36, 2012.

[35] T. Yamashita, K. Yamashita, and R. Kamimura: A stepwise AIC method for variable selection in linear regression, *Communications in Statistics: Theory and Methods*, Vol. 36, No. 13, pp. 2395–2403, 2007.

[36] R. Ho: *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*, Chapman and Hall/CRC, 1996.

[37] Y. Kano and A. Miura: *Graphical Multivariate Analysis by AMOS, EQS and CALIS*, Gendai-Sugaku-Sha, 2002 (in Japanese).

[38] W. Youyou, M. Kosinski, and D. Stillwell: Computer-based personality judgements are more accurate than those made by humans, *Proceedings of the National Academy of Science*, Vol. 112, No. 4, pp. 1036–1040, 2015.

[39] B. Zhu and D. Kaber: Effects of etiquette strategy on human-robot interaction in a simulated medicine delivery task, *Journal of Intelligent Service Robotics*, Vol. 5, pp. 199–210, 2012.

[40] M. Salem, M. Ziadee, and M. Sakr: Effects of politeness and interaction context on perception and experience of HRI, *Proc. of the 5th International Conference on Social Robotics*, pp. 531–541, 2013.

[41] B. Reeves and C. Nass: *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, CSLI Publications, 1996.

---

**Noriko Suzuki**

She received her B.E. and M.E. degrees from Yokohama National University, Japan, in 1992 and 1994, respectively. She received her Ph.D. degree from Nagoya University in 2003. Between 1994–2011, she worked as a researcher at ATR in Japan and during 2006–2012, she worked as an expert researcher at NICT in Japan. She was a post-doctoral fellow at Doshisha University with the support of funding by JSPS during 2012–2013. In 2013, she joined Tezukayama University, where she is currently a lecturer in the Faculty of Business Administration. Her research interests include the design of interaction among people.

**Haruka Shoda**

He received his B.A. degree from Osaka University in 2008. He received his M.A. and Ph.D. degrees from Hokkaido University in 2010 and 2013, respectively. He was a post-doctoral fellow at Doshisha University and Royal College of Music in London with the support of funding by JSPS. He is currently a post-doctoral fellow at Ritsumeikan University, exploring possible connections between music and human well-being.

---

**Rina Matsubayashi**

She received her B.A. degree from Doshisha University in 2017. She is currently a system engineer at SCSK corporation in Japan.

**Mamiko Sakata**

She received her B.S., M.S., and Ph.D. degrees from Kobe University, Japan, in 1997, 1999, and 2002, respectively. In 2005, she joined Doshisha University, where she is currently an associate professor in the Faculty of Culture and Information Science. Her research interests include dynamic structure of multimodal interaction, a learning system of Japanese traditional performing arts, and body culture.