IEEE Robotics and Automation Letters (RAL) paper presented at the
2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Person-directed pointing gestures and inter-personal relationship: Expression of politeness to friendliness by android robots *

Carlos T. Ishi, Ryusuke Mikata, Hiroshi Ishiguro *Member, IEEE*

***Abstract*—Pointing at a person is usually deemed to be impolite. However, several different forms of person-directed pointing gestures commonly appear in casual dialogue interactions. In this study, we first analyzed pointing gestures in human-human dialogue interactions and observed different trends in the use of gesture types, based on the inter-personal relationships between dialogue partners. Then we conducted multiple subjective experiments by systematically creating behaviors in an android robot to investigate the effects of different types of pointing gestures on the impressions of its behaviors. Several factors were included: pointing gesture motion types (hand shapes, such as an open palm or an extended index finger, hand orientation, and motion direction), language types (formal or colloquial), gesture speeds, and gesture hold duration. Our evaluation results indicated that impressions of polite or casual are affected by the analyzed factors, and a behavior's appropriateness depends on the inter-personal relationship with the dialogue partner.**

I. Introduction

In human-human interactions, people exchange non-verbal information such as eye gaze, facial expressions, and hand gestures in addition to linguistic information, and both verbal and non-verbal expressions may change based on the attitudes and relationships among different dialogue partners [1]-[2]. Thus, communication robots that interact with people must be able to express such non-verbal information in a situation-dependent manner.

Numerous studies have been conducted on non-verbal expression in robots [3]. In particular, many studies have focused on generating hand and body movements in robots associated with their speech utterances, in a bid to increase human-likeness and improve the quality of human-robot interaction [4]-[7]. However, most current communication robots are not able to generate behaviors adapted to a situation or an interaction partner. For example, a robot is expected to behave politely when meeting or greeting someone for the first time; however, behaving overly politely to a very close friend may create a cold, distant impression. To forge more intimate human-robot relationships, robots should serve in a more friendly or casual manner, which we believe is the key for strengthening bonds with conversation partners.

Among many non-verbal expressions, we focus on pointing (or deictic) gestures in this study, since they often occur during dialogue interactions and because the way they are expressed affects impressions of politeness or friendliness. For example, pointing with an index finger at a dialogue partner is usually impolite [8]. However, when such person-directed pointing gestures occur in dialogue interactions among friends, these gestures are not particularly interpreted as impolite [9].

No previous studies have clarified how the various factors involved in the different realizations of person-directed pointing gestures affect people's impressions of robot's behaviors. Thus, this study aims on investigating the effects of several factors related with person-directed pointing gestures by accounting for inter-personal relationships. We analyzed human-human dialogue data and designed video-based subjective experiments to evaluate the effects of different pointing gestures by an android robot. We also investigated what kind of behavior is recognized as most appropriate, depending on the relationship with the interaction partner.

This manuscript is organized as follows. In Section II, related works are presented. Section III presents analysis results on person-directed pointing gestures in human-human dialogue interactions. Section IV evaluates our subjective experiments with an android robot for investigating the effects of several factors involved in pointing gestures, including language type, hand shape, hand orientation, motion direction, gesture speed, and hold duration, on the impressions of behavior politeness and interlocutor appropriateness. Discussion and summary of the results are presented in Sections V and VI.

II. Related work

Many studies have addressed non-verbal expressions by robots or agents. In this section, we focus on related studies on pointing gestures and polite expressions, which are the main targets of this study.

The effects of expressions of politeness have been investigated in task-oriented interactions using a small humanoid robot by generating motion types such as bowing and lying [10]. The different behaviors expressed by the robot affected the participants' impressions of it as well as their behaviors toward it. However, pointing gestures were not evaluated.

Regarding studies on pointing (deictic) gestures in human robot interaction, in referential communication, incorporating robot gestures that are accompanied by speech allows humans to more clearly identify target objects [11] and improves their impressions of the robot [4]. Robot pointing gestures in a narrative performance scenario also increase information recall in users [5].

This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas JP20H05576, and JST ERATO Grant Number JPMJER1401.

[1] C. T. Ishi is with the Hiroshi Ishiguro Labs., ATR, Kyoto, Japan (e-mail: carlos@atr.jp), and the Interactive Robot Research Team, Robotics Project, RIKEN, Kyoto, Japan (e-mail: carlos.ishi@riken.jp)

[2] R. Mikata and H. Ishiguro are with the Hiroshi Ishiguro Labs., ATR, Kyoto, Japan (e-mail: ishiguro@irl.sys.es.osaka-u.ac.jp).

In the above studies, the pointing gestures mainly indicated objects or directions without being specifically directed at a dialogue partner. Different models have been proposed for pointing gestures toward an object or a person. Precise pointing gestures were judged to be more acceptable toward an object, but not appropriate toward a person [8]. In another study, politeness behavior through speech and gestures has been evaluated in a robot for guidance and recommendation tasks in a healthcare service setting [12]. Among the gestures, a slow bow is associated with higher politeness levels; a fast pointing action is associated with lower levels.

The above studies indicate that pointing at a person is generally impolite. However, different pointing gesture types were not compared. Perhaps more importantly, since the above situations were limited to service tasks, the relationships between the robot and the user are obviously not very close.

The effects of gesture speed and gesture hold durations also affect impressions of the robot's behaviors [13-14], although they were disregarded in previous studies on pointing gestures.

Thus, in this study, we analyzed various realizations of person-directed pointing gestures and evaluated how different factors influence the impressions of politeness by considering inter-personal relationships.

III. Analysis of Pointing Gestures in Dialogue Interactions

In this section, we analyze person-directed pointing gestures in human-human dialogue interactions (HHI) to verify how different pointing gestures are related to inter-personal relationships.

A. Dialogue dataset and motion annotation

For analysis, we used a dataset of a multimodal three-party conversational speech database collected at our research institute (ATR) [6]. It contains multiple sessions of face-to-face conversations among three speakers. Audio, video, and motion data are available for each speaker. Each dialogue session includes 15 to 40 minutes of free-topic conversations.

We analyzed the data of 7 Japanese dialogue sessions which included a high occurrence of gestures from 13 speakers (7 females and 6 males). The female speakers are research assistants in their 30s and 40s; the male speakers are graduate students in their 20s. Most of the speakers knew each other. In three of the sessions, one pair of speakers met for the first time, while the other two pairs already knew each other. Since the speakers were instructed to talk about any topic, they ended up talking about lunch, work, travel abroad, future plans, etc. For the dialogues between those who were meeting for the first time, topics often focused on mutual acquaintances. The three-party conversations in the dataset included the following three combinations: 2 students + 1 research assistant, 1 student + 2 research assistants, or 3 research assistants. Note that since the coverage of the speakers' ages and inter-personal relationships is limited to the available data, some bias may be added in the analysis results.

For the above dataset, we extracted pointing gesture events and annotated the hand shape, orientation, and motion direction. We manually identified them based on video and hand-motion trajectory displays and only used the person-directed pointing gestures for analysis. About 300 pointing gesture events were extracted from the database. Note that all of these pointing gestures were unconsciously/spontaneously produced by the speakers during the dialogue interactions; they were not coaxed to produce specific gestures and they did not know gestures were being analyzed. The dialogue context was necessary to identify whether a pointing gesture was directed at a person.

Then for each pointing gesture event, the hand shape, orientation, and motion direction were annotated based on the following criteria, which were determined after preliminary observations on different pointing gestures in the database:

- Hand shape: an open palm with all fingers extended (palm), only the index finger extended (index), or other hand shapes, such as only the thumb extended (others).
- Hand orientation: the palm is turned upward (up), downward (down), or to the side or vertically (side).
- Hand motion direction: hand/arm moving from a high position downward (downward or dw), from a low position directly in a forward direction (forward or fw), or others.

The hand shape and orientation categories cover the variations in pointing gestures from previous studies [15-16], while the motion direction categories were newly introduced in this study.

Two research assistants annotated the above labels. Their inter-rate agreements in terms of Cohen's kappa were 0.84, 0.72, and 0.61 for the hand shape, the hand orientation, and motion direction, respectively.

Figure 1 shows the distributions of the pointing gesture types in the dataset. The numbers in parentheses indicate the occurrences for each category. We omitted the "others" categories since their occurrence rates were less than 3%. Fig. 1 shows that index fingers faced the side direction and moving to the forward direction appeared with the highest frequency, and open palms in the upward direction and moving to the forward or downward directions appeared with the second highest frequency.

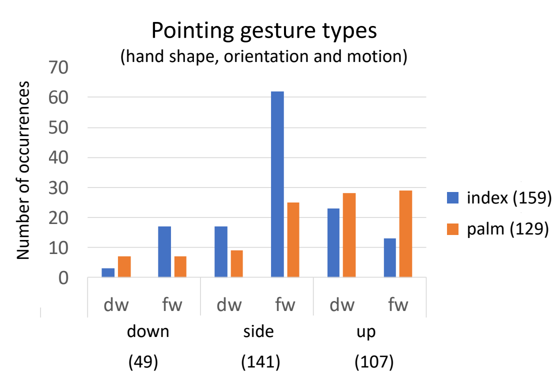

Figure 1. Distributions for different types of pointing gestures in our dataset: hand shape (index or palm); hand orientation (up, side, or down); hand motion (forward or downward)

B. Inter-personal relationship analysis

We analyzed how different pointing gestures occur among various inter-personal relationships between the dialogue partners.

The levels of intimacy and the relative ages were taken into account for categorizing the inter-personal relationships.

- Level of intimacy: first-meeting (when the dialogue partners met for the first time), classmates (of the same university) and colleagues (of the same company).
- Age: younger, about same age or older.

Figures 2 and 3 show the distributions of pointing gesture types for different inter-personal relationship categories. The numbers in parentheses indicate the occurrences for each category. Chi-square tests indicated significant relationships between types of pointing gestures and inter-personal relationship categories (χ^2 (15) = 84.570, $p < 0.01$, for hand shape and motion direction, and $\chi^2(25) = 142.425$, $p < 0.01$, for hand shape and orientation). The symbols in Figs. 2 and 3 represent the gesture types that have significantly higher occurrence rates (* $p < 0.05$, ** $p < 0.01$).

The results in Figs. 2 and 3 show that the index finger is used most frequently at classmates and colleagues of roughly the same age. Open palms are more frequently used on first meetings with both younger and older people as well as with older colleagues. Regarding the motion direction, a forward motion appears with higher occurrence rates for all inter-personal relationship categories, and a downward motion appears with relatively high frequency toward younger dialogue partners.

Another observation is that pointing at older dialogue partners is done less frequently than for younger ones. This

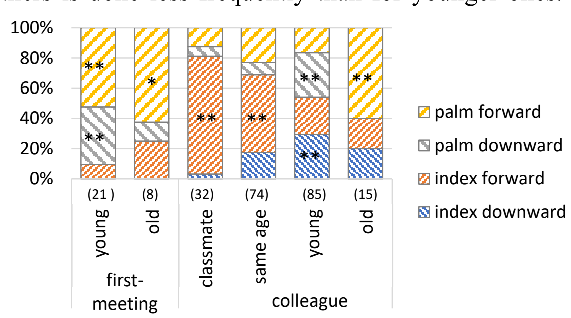

Figure 2. Distributions of pointing gestures types (hand shape and motion direction) for different inter-personal relationship categories (* $p < 0.05$, ** $p < 0.01$)

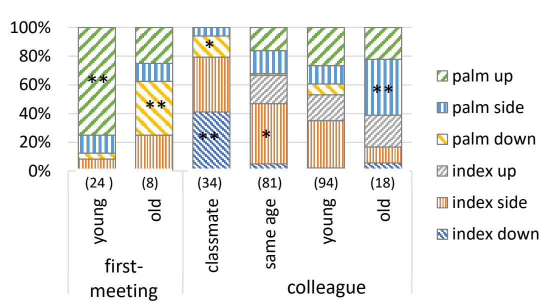

Figure 3. Distributions of pointing gestures types (hand shape and orientation) for different inter-personal relationship categories (* $p < 0.05$, ** $p < 0.01$)

probably reflects the fact that pointing might be perceived as impolite toward an older person, while the reverse is thought to be more acceptable.

IV. Subjective evaluation of pointing gestures by an android robot

The analysis results on human-human dialogue interaction data in Section III suggest that the distributions of pointing gesture motion types were different depending on the inter-personal relationships, meaning that the appropriateness of a behavior may depend on the dialogue partner. Therefore, when designing behaviors in a humanoid robot, it is important to clarify how the several factors involved in different pointing gestures affect the impressions of the robot's behaviors.

In addition to the differences in hand shape, orientation and movement directions during the pointing gestures that are considered in the HHI analyses, the formality expressed by linguistic information such as formal and colloquial, as well as the gesture speed and hold durations also affect the impressions of the robot's behaviors.

In this study, we designed three video-based within-subject experiments using an android robot to clarify how the above factors affect the impressions of the attitudes of politeness as well as their appropriateness toward interlocutors with different inter-personal relationships. Section IV-A describes the general procedure common to all the subjective experiments, and Sections IV-B to IV-D describe the evaluation results for different modalities.

A. General procedure: generation of pointing gestures in an android robot

We used ERICA [17], a female-type android, in all the experiments. Her current version has 44 air actuators: 13 DOFs for her face, 3 DOFs for her head motion, 1 DOF for the base of her neck, 3 DOFs for her torso motion, and 12 DOFs for each hand/arm: 2 for her shoulder, 3 for her upper arm, 2 for her forearm, 2 for her wrist, 1 for her thumb, 1 for her index finger, and 1 for her remaining three fingers.

To evaluate the effects of different pointing gestures, we created gesture motions for ERICA by manually editing the hand/arm actuator values. Our research group members (including research assistants with experience on gesture annotation) preliminarily checked these pointing gesture motions to verify whether they are in synchrony with the utterances and look sufficiently humanlike. Although the human likeness of the created motions is constrained by hardware limitations on the motion speed and hand shape (e.g., the fingers cannot be completely folded during an index finger pointing gesture), the motions contain adequate quality for expressing the desired gesture motions.

For the gesture phase durations, we set the default values for the gesture preparation, hold, and retraction phases to 0.8, 1.0, and 1.4 seconds, respectively, which are the same values used in previous studies on hand gesture generation [6]. The durations of the preparation and retraction phases are slightly longer than the average durations of human data (0.7 and 0.9 s) to avoid jerky movements [6].

Regarding the actuators other than those for the hands/arms, we applied the methods proposed in previous studies on speech-driven lip and head motion generation [18-20]. Since the facial and head actuators are independent from the hand/arm actuators, these two streams (the automatically generated facial and head motion data and the manually generated hand/arm motion data) are simply concatenated to generate the robot actuator commands.

Video clips of the android's motions were recorded for each of the target conditions evaluated in the subjective experiments, which were designed in a within-subject manner. The participants watched all of the video clips and answered a questionnaire about their impressions of the robot's behaviors of each video clip. None of our participants were involved in robotics research. The videos show only the robot (i.e., not the interlocutor), so the participants evaluated the robot's behavior from a third-person perspective. Details of the target conditions and procedures will be described for each experiment in the following sections.

B. Experiment 1: Effect of pointing gesture type (shapes and motions) and language type

Our first experiment investigated the effects of the pointing gesture types in terms of hand shape and hand/arm movement, and language type (degree of formality).

1) Experimental setup: The independent variables are the pointing gesture type, including combinations of hand shape (open palm or index finger), hand/arm movement direction (forward or downward), and the language type (formal or colloquial). The measurements are the perceived degree of the behavior politeness and the interlocutor appropriateness. We made the following hypotheses, based on the HHI analysis results that show higher occurrences of specific motion types depending on the inter-personal relationship.

Hypothesis 1a: the hand shape, the motion direction, and the language type affect the impression of politeness; open palm, downward motions, and formal language will be perceived as more polite than their counterparts (index finger, forward motions, and colloquial language).

Hypothesis 1b: a behavior's appropriateness is affected by inter-personal relationships; polite behaviors are more appropriate toward interlocutors with distant relationships, and casual behaviors are more appropriate toward interlocutors with close relationships.

For the types of pointing gestures, we created the following four motion types by combining the hand shape (open palm or index finger) and the movement direction (downward or forward). Fig. 4 shows snapshots of the created motion sequences for two of the gesture types:

- *palm downward*: moving the hand in a downward direction from a higher position to a pointing position with an open palm up. This motion is expected to be the most polite pointing gesture.
- *palm forward*: directly moving the hand in a forward direction from a resting position to a pointing position with the palm up. This motion is expected to look slightly more casual than the palm downward.
- *index downward*: this motion is identical to the palm downward with an index finger pointing at the interlocutor.

Figure 4. Snapshots of motion sequences for "palm downward" (left panels) and "index forward" (right panels).

- *index forward*: this motion is identical to the palm forward with an index finger pointing at the interlocutor. This motion is expected to be appropriate/acceptable in casual situations.

While making the robot motions for the index finger, we believed that pointing with an index finger at an interlocutor will be construed as impolite. However, pointing up from a lower height at her/him will feel less impolite if the interpersonal relationship between the speakers is relatively close. Indeed, more than half of the pointing gestures were achieved using an index finger pointing at an interlocutor in friendly dialogue interactions among people who have close relationships.

For the language that accompanied the pointing gestures, we considered two styles: formal and colloquial. Analysis of human-human dialogue interaction data has suggested that pointing at an interlocutor often occurs during dialogue interactions when asking a question or stating a request to emphasize that it is now the interlocutor's turn (about 45% of the pointing gestures in the analyzed dataset). We accounted for this situation by creating two language types for the same question. One is a formal sentence: "*anata wa aki wa suki desuka*?"; the other is colloquial: "*de, aki wa suki nano?*" Both sentences mean "do you like autumn?"

For the formal sentence ("*anata wa aki wa suki desuka?*") we adjusted the timing of its pointing to the phrase "*anata wa*" ("you"). In the colloquial sentence, the subject "you" is omitted, so that the pointing timing was adjusted at the end of the sentence around "*suki nano*" ("like?"). We also tested the pointing's timing at the end of the formal sentence, but our preliminary evaluation indicated that synchronization with "*anata*" ("you") was preferred.

We recorded video clips ranging from around 5 to 10 seconds for all of the utterance and motion types. To investigate the effects of the pointing gesture types and the language types, two sets of four videos were prepared. Each set was composed of four motion types: palm downward, palm forward, index downward, and index forward.

For training before the main experiment, the participants watched a sample video (palm downward motion with formal language) and answered a questionnaire, which included several questions, including subjective impressions on politeness and the appropriateness of the behavior for interlocutors with different inter-personal relationships: first meeting or close friend; older, younger, or same age. The politeness impressions were graded on a 7-point scale (-3 for casual/careless, 0 for too difficult to decide, 3 for polite/careful), and the appropriateness was judged for each interlocutor type in a checkbox style (1 for appropriate and 0 for inappropriate).

In the main experiment, we evaluated each set of four videos. Within each one, the video order was randomized among the different motion types. The participants first watched all four videos and then answered questionnaires for each video after watching each one again. In the main evaluation, each video could be seen twice at most. This procedure was repeated for the other set of four videos. Thus, in this first experiment, participants evaluated 8 videos: 2 sets x 4 motion types.

Thirty-six subjects (11 males and 25 females, whose ages ranged from 20s to 40s, mean, 28.7, standard age deviation, 11.2) participated in the experiments. None had any previous involvement in robotics research.

2) Experimental results: Fig. 5 shows the subjective evaluation results for the appropriateness (7 point-scale from casual/careless to polite/careful) of the four different pointing gesture types and the two language types (formal and colloquial). The vertical axis represents the perceptual scores.

Statistical significance tests were conducted on the subjective scores for politeness through a two-way repeated measures ANOVA with language type (formal or colloquial), and motion type (combination of hand shape: palm or index, and motion direction: downward or forward) as independent factors. No significant two-way interaction was found ($F(3,105) = 2.2$, $p = 0.09$, $\eta_p^2 = 0.06$). Main effects were significant in both language types ($F(1,35) = 20.9$, $p < 0.01$, $\eta_p^2 = 0.37$) and in the motion types ($F(3,105) = 130.9$, $p < 0.01$, $\eta_p^2 = 0.79$).

The results in Fig. 5 show that overall, the index-finger-based pointing motions were evaluated as casual, and the palm-based pointing motions were evaluated as polite ($p < 0.01$). Colloquial sentences were rated as less polite than their formal counterparts for all pointing gestures ($p < 0.05$), except for index finger forward, which is the most casual behavior. No significant differences were found between the downward and forward movements regarding polite impressions. These results confirm Hypothesis 1a for the hand shape and language types, but not for the motion direction.

Figure 6 shows the subjective evaluation results of interlocutor appropriateness by different pointing gestures and language formality types. The vertical axis represents the average scores of the participants who evaluated whether a motion is appropriate for a specific interlocutor (first-meeting same age, first-meeting younger, first-meeting older, close friend, close older).

Chi-square tests identified a significant relationship between pointing gesture types and interlocutor categories for appropriateness ($\chi^2(28) = 170.91$, $p < 0.01$). Symbols "^" and "v" in Fig. 6 indicate significantly higher and lower rates for behavior appropriateness toward different interlocutors.

The results in Fig. 6 indicate that for first-meeting older interlocutor, palm-based pointing gestures with formal language are overwhelmingly judged to be appropriate (with more than 80% of the votes), and all other pointing gesture types are inappropriate (with less than 10% of the votes). For

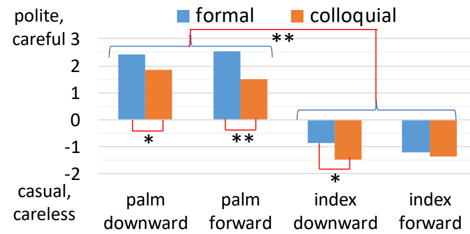

Figure 5. Subjective evaluation results for polite impressions by different pointing gestures (hand shape and motion direction) and language types (*$p < 0.05$, ** $p < 0.01$)

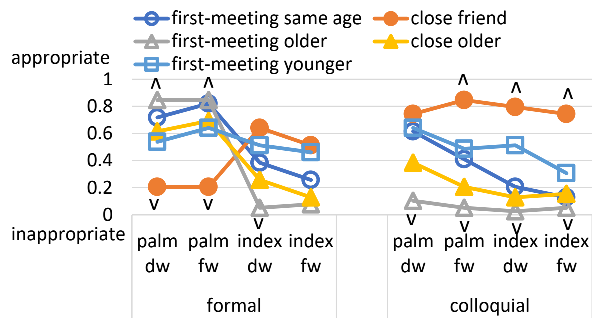

Figure 6. Subjective evaluation results for interlocutor appropriateness by different pointing gesture types and language types ("^" and "v" indicate significantly higher and lower, respectively, with $p < 0.05$ levels)

close friends, a reverse trend was observed, where index-based pointing gestures with colloquial sentences are judged to be appropriate, and palm-based gestures with formal language are judged as inappropriate (with less than 20% of the votes). The results for the other interlocutor categories, which were between first-meeting older person and close friend, are ranked as follows: first-meeting older, close older, first-meeting same age, first-meeting younger, and close friend. The above results confirm Hypothesis 1b, which predicted that behavior appropriateness will be affected by inter-personal relationships.

C. Experiment 2: Effects of gesture speed and hold duration

The second part of our experiment investigated the effects of gesture motion speed and gesture hold durations on behavior appropriateness impressions. The same subjects from Experiment 1 participated in this experiment.

1) Experimental setup: Two sets of motions were prepared for different speed and hold durations. The independent variables are the pointing gesture types (including combinations of language and motion types) and gesture speed for the first set of motions and pointing gesture types and gesture holding durations for the second set of motions. The measurement is the behavior politeness level. The following are our hypotheses, which are based on past evidences that motion speed and gesture hold durations affect the impressions of robot's attitudes [13-14]:

Hypothesis 2a: the gesture speed affects the impressions of politeness; faster motions will look more casual, and slower motions will look more polite.

Hypothesis 2b: the gesture hold duration affects impressions of politeness; short hold durations will look more casual, and longer hold durations will look more polite.

For the pointing gesture type in this experiment, we selected three combinations of language and motion types. The first is a formal sentence with a palm downward motion, which expresses the most polite manner. The second is a colloquial sentence expressed in an index finger forward motion, which expresses the most casual manner. The third is a colloquial sentence with a palm forward motion, which expresses an intermediate impression between the first two.

We evaluated the effects of gesture speed by changing the duration of the preparation and retraction phases in seven scales: {0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0} times the default durations (0.8 seconds for preparation and 1.4 seconds for retraction) used in Experiment 1 (Section IV). Thus, scales smaller than 1.0 correspond to faster motions; scales larger than 1.0 correspond to slower motions.

In the same way, for evaluating the effects of gesture holding, the duration of the hold phase was changed in five scales: {0.0, 0.5, 1.0, 1.5, 2.0} times the default duration (1 second) used in Experiment 1. Thus, scales smaller than 1.0 correspond to shorter holding durations; scales larger than 1.0 correspond to longer holding durations.

Motion sets for different scales of gesture speed and gesture hold duration were created for each of three combinations: {*palm downward formal*, *palm forward colloquial* and *index forward colloquial*}.

In this experiment, each motion set was evaluated, in a way that the video order sequence was kept among different scales. To alleviate the fatigue of the participants, the experiments were conducted by the following procedure. They first watched all the videos in a particular motion set. Then they watched the default motion (1.0) and in sequence the motions with scales smaller than 1.0, until they noticed a change in their impressions. Then they answered the same questionnaire sheet for Experiment 1 on the video for which they noticed a change and filled out the corresponding video number. If they did not notice any change, they did not answer the questionnaire. The same procedure was conducted for the videos with scales exceeding 1.0. This entire procedure was repeated for the other two motion sets. In this way, the participants answered at most 18 questionnaires (3 sets x (3 speed + 3 hold) = 18 conditions).

Twenty-six of the thirty-six subjects in Experiment 1 participated in Experiment 2 (7 males and 19 females, mean, 26.9, standard deviation, 10.0 years). The evaluations of both Experiments 1 and 2 took from 1 to 1.5 hours.

2) Experimental results: Fig. 7 shows the subjective evaluation results of the gesture speed and hold duration on the polite impressions for the three combinations of motion and language types.

We conducted two-way repeated measures ANOVAs on the motion type and gesture speed categories: fast, default, and slow. Main effects were found in both motion type ($F(2,50) = 48.9$; $p < 0.001$; $\eta_p^2 = 0.66$) and motion speed ($F(2,50) = 8.7$; $p < 0.001$; $\eta_p^2 = 0.26$). No mutual interaction was found between them ($F(4, 100) = 0.88$; $p = 0.48$; $\eta_p^2 = 0.03$).

Two-way repeated measures ANOVAs were also conducted on the motion type and gesture hold duration categories (short, default, and long). Main effects were found in both the motion type ($F(2,50) = 42.3$; $p < 0.001$; $\eta_p^2 = 0.63$) and the hold duration ($F(2,50) = 6.9$; $p < 0.005$; $\eta_p^2 = 0.22$). Mutual interaction between motion type and hold duration was also found ($F(4,25) = 4.9$; $p = 0.001$; $\eta_p^2 = 0.17$). Multiple comparisons (through Ryan's method) verified the significant differences between specific motion pairs.

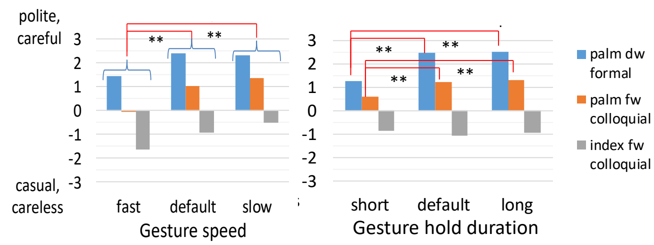

Figure 7. Subjective evaluation results of gesture speed and hold duration on carefulness impression, for three combinations of motion and sentences types (** $p < 0.01$).

Overall, faster motion and shorter hold durations tend to be perceived as less polite than the default and slow speed ($p < 0.01$) and the default and longer hold durations ($p < 0.01$), respectively. This was not true only for the index forward colloquial motion type (i.e., colloquial language while pointing forward with an index finger). In this case the behaviors were all judged as casual, regardless of the hold duration.

Although slow motions and longer hold durations were judged to be more polite than the fast motions ($p < 0.01$), they did not significantly increase the impressions of politeness, relative to the default motions.

The following are the percentages of participants who did not feel changes in their impressions of the robot's behavior: 16% for the fast motions, 15% for the slow motions, 30% for the short hold durations, and 33% for the longer hold durations.

Regarding the speed scale, the average scales for the fast motions were around 0.57 times the default values which correspond to around 0.5 and 0.8 seconds for the preparation and retraction durations. Those for the slow motions were around 1.62 times the default values which correspond to around 1.3 and 2.2 seconds for the preparation and retraction durations.

For the hold duration scale, the average scales for the short motions were around 0.15, which correspond to around 150 ms, and those for the long motions were around 1.84, which correspond to around 1.8 seconds.

D. Experiment 3: Effects of hand orientation

The evaluation results in Experiment 1 show clear differences between hand shape (open palm or index finger) and slight but not significant differences in the impressions of different hand/arm movements (downward or forward). Our third experiment investigated the effects of hand orientation on impressions of behavior politeness.

1) Experimental setup: The independent variables are the pointing gesture type (including combinations of hand shape and hand orientation types) and the language type. The measurement is the behavior politeness level. The following hypothesis is based on the HHI analysis results where hands that face up tend to appear with higher frequencies toward people who are meeting for the first time, and hands that face to the side appear with higher frequencies toward closer interlocutors.

Hypothesis 3: the hand orientation affects the impressions of politeness; a hand facing up will be felt as more polite than a hand that faces to the side.

For this experiment, we used combinations of hand shape (open palm or index finger), hand orientation (hand facing up or to the side), and language type (formal or colloquial), resulting in eight motion types. Fig. 8 shows snapshots of the hand shapes created for the android. The motion direction was fixed down for the formal and forward for colloquial language.

The procedure conducted for the subjective evaluation of the behavior impressions resembled that in Experiment 1. Although ERICA's visual appearance (clothes and hair style) in Experiment 3 is different from that in Experiments 1 and 2, we don't believe this will not strongly affect the impressions of the generated behaviors.

In Experiment 3, we investigated a situation where the interlocutor's voice is also played to simplify understanding of the dialogue context:

Formal language: Interlocutor: "*Hai, sangatsu de sotsugyou desune*" ("*Yes, I'm graduating this March.*"); ERICA: "*Sotsuron donna kanji desuka?*" ("*How is your bachelor thesis going?*"; Interlocutor: "*Kekkou yabai desu.*" ("*Not very well. . .* ")

Casual language: Interlocutor: "*Un, sotsugyou shichau*"; ERICA: "*Sotsuron donna kanji?*"; Interlocutor: "*Kekkou yabai.*" (identical meaning as the formal language).

Figure 8. Snapshots of created hand shapes and orientations: palm up, palm to the side, index finger up, and index finger to the side

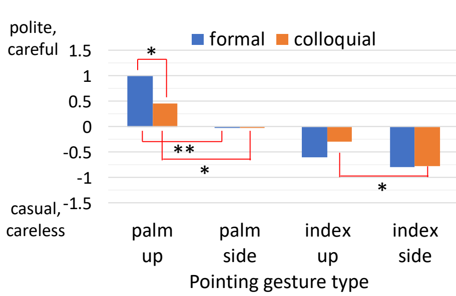

Figure 9. Subjective evaluation results for politeness impressions by different pointing gesture types (hand shape and orientation) and sentences types (* $p < 0.05$, ** $p < 0.01$. Significance symbols between palm and index are omitted, but are all significant with $p < 0.05$).

Thirty-one undergraduate students (23 males and 8 females, all around 20 years old) with no robotics experience participated in the experiment. They are different from the participants of the previous experiments.

2) Experimental results: Fig. 9 shows the subjective evaluation results of the politeness impressions for different language and gesture types.

Two-way repeated measures ANOVA were conducted on the subjective scores for politeness with language type (formal and colloquial) and motion type (hand shape: palm or index finger) vs. hand orientation (up or to the side)) as independent factors. Two-way interaction was found to be significant ($F(3,90) = 2.81$; $p < 0.05$; $\eta_p^2 = 0.08$). Main effects were significant in the gesture type ($F(3,90) = 18.35$, $p < 0.001$; $\eta_p^2 = 0.38$), but not in the sentence type ($F(1,30) = 0.19$; $p = 0.65$; $\eta_p^2 = 0.007$). Multiple comparisons (through Ryan's method) also verified the significant differences in specific motion pairs.

For the hand shape results, palm gestures were significantly more polite than index ones ($p < 0.05$). Regarding the hand orientation categories, hands facing up were significantly more polite than hands to the side in the formal palm ($p < 0.01$), colloquial palm, and colloquial index types ($p < 0.05$). These results confirm Hypothesis 3, except for the formal index type.

V. Discussion

The results of Experiment 1 (Section IV-B) indicated that hand shape and language type significantly affected the impressions of behavior politeness, but not the motion direction. This implies that when controlling the politeness levels of the robot's behaviors, the effects of motion direction are less relevant than the other factors.

As for the results of Experiment 2 (Section IV-C), faster motions and shorter hold durations decreased impressions of politeness. However, slower motions and longer hold durations did not significantly increase them, relative to the default motions. This suggests that when controlling the politeness levels of the robot's behaviors, more speed and shorter hold durations will reduce such levels. Slower motion and longer hold durations may be more related to the speaker's personality than politeness.

The results in Experiment 3 (Section IV-D) indicated that the politeness impressions are higher for hands that face up than to the side, except for the index finger with formal sentences (formal index). This suggests that the hand orientation effects are stronger in pointing gestures using an open palm.

Our overall results for inter-personal relationships suggest that casual behaviors are appropriate or acceptable with dialogue partners who share close relationships (friends of about the same age), and polite behaviors are appropriate with dialogue partners who share more distant relationships (first-meeting and older person). Note that overly polite behaviors were judged as inappropriate with close friends of the same age, since they would probably be felt as cold or distant. Therefore, if the robot can identify the relationship with the interlocutor, these distributions can be used to appropriately select its behavior. This is one step of our future work.

The above results on the subjective evaluations roughly match the higher occurrences in the distributions of the different motion types for the various inter-personal relationships analyzed in the human data in Section III. However, some differences surface for behaviors toward first-meeting older interlocutors in the analysis data and in the subjective experiments. In the former, we found a difference of around 10 years between the dialogue partners. But since their statuses were (younger) students and (older) research assistants, the distance between the dialogue partners was relatively trivial. However, one might expect a change in behavior if their relationship were (younger) subordinate and (older) boss. Quantifying the distance between interlocutors (e.g., based on a relationship closeness inventory [21]) for better categorization of inter-personal relationships is another topic to consider for future investigation.

A limitation of the video-based study design described in Section IV is that the subjective grades are based on a third-person perspective. The impressions of the interlocutor appropriateness might be different in face-to-face interactions. A possible way to design such an experiment is teleoperating the robot by both known and unknown interlocutors.

Finally, different languages and cultures might employ other manners to express or feel politeness and friendliness. For example, different pointing gestures have contrastive meanings in Naples (Italy) and in Australian Aboriginal languages [15]. However, it remains unclear how they differ by inter-personal relationships. Undoubtedly, part of the results in the present study are specific to Japanese language and culture, but most of the results are probably applicable for other languages. Nonetheless, the methodology for evaluation can be extended to any language.

VI. Conclusion

We analyzed different pointing gestures in human-human interactions from an inter-personal relationship perspective and investigated how several factors related to pointing gestures expressed by an android robot are perceived in terms of politeness and inter-personal relationships.

Our subjective evaluation results indicated that hand shape has the strongest effect on the impression of behavior politeness (an open palm is more polite than an index finger), followed by hand orientation (hands up are more polite than to the side), and formal language is more polite than colloquial. Evaluation of gesture speed and hold duration indicated that faster motion and shorter hold durations reduce the levels of politeness. Both HHI data analysis and subjective evaluation results using a robot indicated that the choice of a pointing gesture is affected by inter-personal relationships.

These results can be directly used to control the levels of politeness that the robot wishes to express toward specific interlocutors, based on inter-personal relationships. To achieve such results, the robot must be able to estimate its relationship with its dialogue partner. In future work, we will apply these results to dialogue robot systems and evaluate the effects of temporal behavior changes during human-robot interactions.

Acknowledgments

We thank Taeko Murase, Yuka Nakayama, and Kyoko Nakanishi for their contributions on the speech and motion data analyses.

References

[1] [1] L. K. Guerrero and K. Floyd, Nonverbal communication in close relationships. Lawrance Erlbaum Associates, Publishers, 2006.

[2] P. Wagner, Gesture and speech in interaction: An overview. *Speech Communication 57*(2014), 209–232, 2013.

[3] N. Mavridis, A review of verbal and non-verbal human-robot interactive communication. *Robotics and Autonomous Systems 63*, 22–35, 2015.

[4] M. Salem, S. Kopp, I. Wachsmuth, K. Rohlfing, and F. Joublin, Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics 4*(2), 201–217, 2012.

[5] C.M. Huang and B. Mutlu. Modeling and Evaluating Narrative Gestures for Humanlike Robots. *Robotics: Science and Systems*, pp. 57-64, 2013.

[6] C.T. Ishi, D. Machiyashiki, R. Mikata, H. Ishiguro. A speech-driven hand gesture generation method and evaluation in android robots. *IEEE Robotics and Automation Letters 3*(4), 3757–3764, July 2018.

[7] Y. Yoon, W. Ko, M. Jang, J. Lee, J. Kim, G. Lee. "Robots learn social skills: End-to-end learning of co-speech gesture generation for humanoid robots," Proc. *Int. Conf. on Robotics and Automation*, 4303-4309, 2019.

[8] P. Liu, D.F. Glas, T. Kanda, H. Ishiguro, N. Hagita, "It's not polite to point: Generating socially-appropriate deictic behaviors towards people," Proc. *Int. Conf. on Human-Robot Interaction*, 2013, pp. 267–274.

[9] C. Ishi, R. Mikata, H. Ishiguro. "Analysis of the factors involved in person-directed pointing gestures in dialogue speech," Proc. *Int. Conf. on Speech Prosody*, pp. 309–313, May 2020.

[10] T. Nomura and K. Saeki, Effects of polite behaviors expressed by robots: A psychological experiment in Japan. *International Journal of Synthetic Emotions 1*(2), 38–52, 2010.

[11] R. Fang, M. Doering, and J. Y. Chai, "Embodied collaborative referring expression generation in situated human-robot interaction," Proc. *ACM/IEEE Int. Conf. on Human-Robot Interaction*, 2015, pp. 271–278.

[12] N. Lee, J. Kim, E. Kim, and O. Kwon, The influence of politeness behavior on user compliance with social robots in a healthcare service setting. *Int. Journal of Social Robotics 9*(5), 727–743, 2017.

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

[14] C.T. Ishi, R. Mikata, T. Minato, H. Ishiguro. "Online processing for speech-driven gesture motion generation in android robots," Proc. *IEEE-RAS Int. Conf. on Humanoid Robots*, pp. 508–514, Oct. 2019.

[15] S. Kita. Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes 24*(2), 145–167, 2009.

[16] H. Cochet and J. Vauclair. Pointing gesture in young children: *Hand preference and language development. Gesture 10*(2-3), 129–149, 2010.

[17] D.F. Glas, T. Minato, C.T. Ishi, T. Kawahara, and H. Ishiguro, "ERICA: The ERATO Intelligent Conversational Android," Proc. *IEEE Int. Conf. on Robot and Human Interactive Communication*, pp. 22–29, 2016.

[18] C. Ishi, C. Liu, H. Ishiguro, N. Hagita. "Evaluation of formant-based lip motion generation in tele-operated humanoid robots," Proc. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2377–2382, Oct. 2012.

[19] C. Ishi, T. Funayama, T. Minato, and H. Ishiguro. "Motion generation in android robots during laughing speech," Proc. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 3327–3332, 2016.

[20] 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 2*(3), 1748–1754, July 2017.

[21] E. Berscheid, M. Snyder, and A.M. Omoto. The relationship closeness inventory: Assessing the closeness of interpersonal relationships. *Journal of Personality and Social Psychology 57*(5), 792–807, 1989.