



# Analysis of the factors involved in person-directed pointing gestures in dialogue speech

Carlos Ishi<sup>1</sup>, Ryusuke Mikata<sup>1</sup>, Hiroshi Ishiguro<sup>1</sup>

<sup>1</sup>ATR Hiroshi Ishiguro Labs.

carlos@atr.jp, mikata.ryusuke@irl.sys.es.osaka-u.ac.jp, ishiguro@sys.es.osaka-u.ac.jp

## Abstract

Pointing gestures directed to a person are usually taken as an impolite manner. However, such person-directed pointing gestures commonly appear in casual dialogue interactions, and convey important visual prosodic information. In this study, we extracted pointing gestures appearing in a three-party spontaneous dialogue database, and analyzed several factors including gesture type (hand shape, orientation, motion direction), dialogue acts, inter-personal relationship and attitudes. Analysis results indicate that more than half of the observed pointing gestures use the index finger towards the interlocutor, but are not particularly perceived as impolite. Pointing with the index finger moving in the forward direction was found to be predominant towards interlocutors with close relationship, while pointing with the open palm was found to be more frequent towards first-met person or older person. The majority of the pointing gestures were found to be used along with utterances whose contents are related or directed to the pointed person, while part were accompanied with attitudinal expressions such as yielding the turn, attention drawing, sympathizing, and joking/bantering.

**Index Terms:** speech and gesture, pointing gestures, visual prosody, intimacy, attitude, paralinguistic information

## 1. Introduction

The background of this study is the generation of natural motions (i.e., visual prosody) in humanoid robots, matched with the speech utterances. So far, the relations between speech and several modalities including facial, head, body and hand movements have been investigated, accounting for dialogue act functions, and emotion expressions such as laughing speech and surprise utterances [1-4], and several methods for generating natural motions in humanoid robots from text and prosodic features embedded in the speech signal have been proposed [5-8]. Among these modalities, the generation of hand gestures still remains a challenging task, since several types of gestures (including iconic, metaphoric, deictic/pointing, beat) co-occur with speech and are affected by several linguistic and contextual factors [9-10].

Several studies have been conducted on text or speech-driven gesture generation. For example, lexicon-based approaches have been proposed for generating iconic gestures in [11-13], and metaphoric gestures in [14]. Prosodic information has also been exploited when generating hand gestures, mainly by considering relations between prosodic focus (emphasis) and beat gestures [15]. It is reported that apexes of gestural strokes and pitch accents aligned consistently, and gestural phrases and intermediate phrases aligned quite often. The method proposed in [8] uses both text and prosodic information for generating hand gestures. Although natural humanlike hand motions could be generated, it has been observed that pointing gestures are

more difficult to be automatically generated from text and speech only.

Some studies on pointing gestures show that in referential communication the incorporation of pointing gestures into an embodied collaborative agent allow humans to better identify target objects and offer higher communicative accuracy [16, 17]. However, in these studies the displayed pointing gestures are mainly for indicating objects or directions, and are not specifically directed to a dialogue partner.

Although the act of pointing towards a person is usually taken as an impolite manner, such person-directed pointing gestures often appear in dialogue interactions. We consider that different types of pointing gestures may convey different attitudes, such as politeness, friendliness/intimacy, or rudeness. Furthermore, the use of different types of pointing, such as with the open palm or with the index finger, may provide different impressions, depending on the situation and the inter-personal relationship between the dialogue partners. Thus, for a multimodal dialogue interaction, it is important to clarify what are the factors involved in person-directed pointing gestures.

In the present study, we analyzed pointing gesture events in face-to-face human interactions appearing in a multimodal three-party dialogue database, and investigated several factors including gesture type (hand shape, hand orientation, motion direction), dialogue acts, inter-personal relationship and the attitudes conveyed by the pointing gestures.

## 2. Analysis data

### 2.1. Description of the data

For analysis, we use a dataset of the multimodal three-party conversational speech database collected at our research institute (ATR). The database contains multiple sessions of face-to-face conversations among three speakers. The data includes audio, video and motion information of each speaker. Fig. 1 shows a picture of the data collection environment setup.

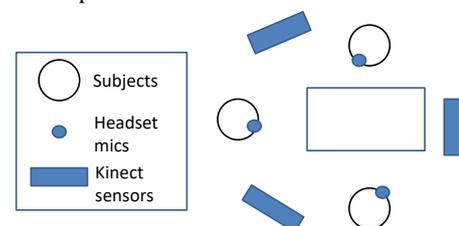


Figure 1. Environment setup for the three-party dialogue collection.

Three headset microphones (DPA4060) and three RGB-D sensors (Microsoft Kinect-V2) were used to capture data of each speaker. The dialogue participants had a seat in chairs around a table, with a distance of about 2 meters between each other. The table is 60cm height, so that the Kinect sensors can detect hand motions around the participants' knee area. The

video data from the Kinect sensors were used in the analysis of the present work.

Each dialogue session comprises about 15 to 30 minutes of random topic conversations. The speech utterances are segmented in phrase units (accentual phrases) and text-transcribed by a native speaker.

For the present analysis, data of 7 Japanese dialogue sessions by a total of 13 speakers (7 female and 6 male speakers, from the Kansai area in Japan) were used. The ages of the female speakers range from 30s to 40s, while all male speakers are graduate students in their 20s. Almost all speakers knew each other. In three of the sessions, one of the speakers met at the first time. The speakers were instructed to talk about any topic, so that resulting daily-life topics such as lunch, work, their experiences abroad, their future plans. For the dialogues between first-met people, topics about a known person in common were predominant.

## 2.2. Annotation data

Several types of annotations were conducted to investigate the factors involved in person-directed pointing gestures.

Firstly, the pointing gesture events were manually identified based on video and hand motion trajectory displays. Only person-directed pointing gestures were identified. About 300 pointing gesture events were extracted from the database. Note that all these pointing gestures were unconsciously (spontaneously) produced by the speakers during the dialogue interactions (i.e., speakers were not instructed to produce specific gestures). The dialogue context was necessary to identify if the pointing gesture was directed to a person.

Then, for each pointing gesture event, the accompanied hand shape, orientation and motion were annotated according to the following criteria:

- Hand shape: open palm with all fingers extended (“palm”), only index finger extended (“index”), other hand shapes (“other”).
- Hand orientation: palm is faced to upward direction (“up”), downward direction (“down”), or side direction (“side”).
- Hand motion direction: hand/arm moving from a high position to the downward direction (“downward” or “dw”), hand/arm moving from a low position straightly to the forward direction (“forward” or “fw”), others.

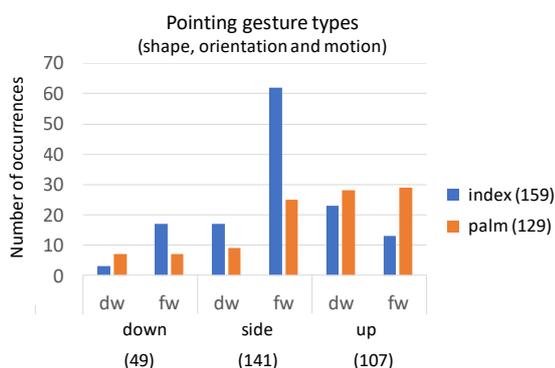


Figure 2. Distributions for different types of pointing gestures in our dataset (hand shape: index or palm; hand orientation: up, side, down; hand motion: forward or downward). The numbers in parentheses indicate the total number of occurrences for each category.

Fig. 2 shows the distributions for the pointing gesture types appearing in the dataset. It can be observed in Fig. 2 that index finger faced to the side direction and moving to the forward direction appeared with the highest frequency, while the open palm faced to the upward direction and moving to forward or downward directions appeared with the second highest frequencies.

Next, the pointing gesture functions were annotated. This is a more difficult task, since annotator has to guess what was the purpose of the speaker on pointing towards the interlocutor. Since there are no previous studies describing the functions of pointing gestures in dialogue interactions, the following categories were established, after discussions among three annotators (including researchers and research assistants with experience in other paralinguistic data annotation).

- “ask” (turn yielding): usually appearing at the end of a question or confirmation (yes-no questions) in order to explicitly yield the turn to a specific dialogue partner. It replaces the phrase “you talk next”.
- “sympathy”: usually co-occurring with backchannel utterances such as “sou!”, “wakaru!” (“right!”, “exactly!”), when the speaker expresses strong agreement or sympathy to the interlocutor’s utterance.
- “joke” (banter): pointing with the attitude of joking/bantering with a dialogue partner. However, the purpose is to soften the atmosphere, and not to offend the dialogue partner.
- “blame”: pointing with the attitude of blaming/criticizing the dialogue partner.
- “look down”: pointing with the attitude of looking down the dialogue partner.
- “assert”: pointing accompanied with a strong assertion by the speaker.
- “main”: pointing towards the main dialogue partner (i.e., the person whom the current utterance is directed to) with the purpose of clarifying or emphasizing that the current topic is related to that person.
- “sub”: pointing towards a third person (who is present in the dialogue space but is not the main interlocutor the speaker is interacting), with the purpose of clarifying or emphasizing that the current topic is related to that person.
- “attention”: usually appearing at the beginning of an utterance in order to draw attention from the dialogue partner. (This item was included as a sub-category of “main”.)

Fig. 3 shows the distributions for the pointing gesture functions. The “main” and “sub” may co-occur with other categories, but the occurrence rates for “main” and “sub” shown in Fig. 3 are excluding co-occurrences.

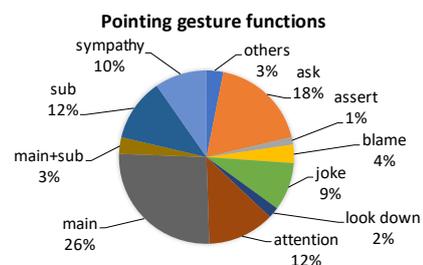


Figure 3. Distributions of different functions carried by pointing gestures found in the dataset.

The “main” category appeared with the highest frequency, followed by “ask”, “attention”, “sympathy” and “joke”. Other categories such as “provoking” and “threatening” were also taken as candidates, but these appeared with very few or no occurrences in this dataset.

The situation in which a pointing gesture occurred was also annotated. The following categories were found:

- situation: “excited”, “funny”, “joyful”, “serious” and “normal”.

Other situations such as “sad” and “angry” were also taken into account, but these did not appear in the current dataset. Fig. 4 shows the distributions of different situations where a pointing gesture occurred.

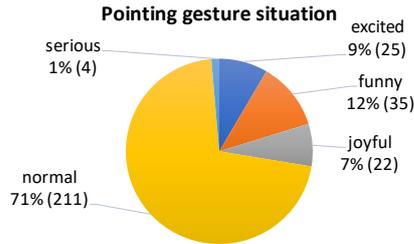


Figure 4. Distributions of different situations for the pointing gestures found in the dataset.

From the distributions in Fig. 4 it can be observed that 70% of the pointing gestures occurred in normal situations, i.e. when no specific emotional expression is accompanied with the pointing gestures.

### 3. Analysis results

We analyzed how different gesture types occur in relation to the inter-personal relationship between the dialogue partners.

The level of intimacy and the ages were taken into account for categorizing inter-personal relationship.

- Level of intimacy: “first-met” when the dialogue partners met as the first time, “classmate” and “colleagues”.
- Age/generation: “younger”, “same age” (including same school grade or same generation) and “older”.

Figs. 5 and 6 show the distributions of pointing gesture types for different inter-personal relationship categories.

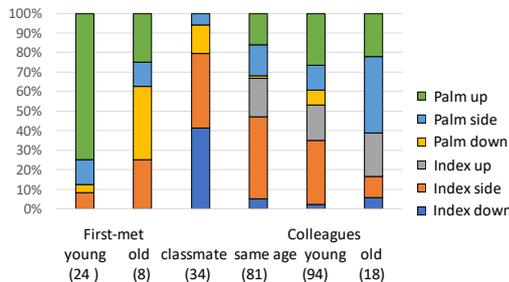


Figure 5. Pointing gestures types (hand shape and orientation) and inter-personal relationship

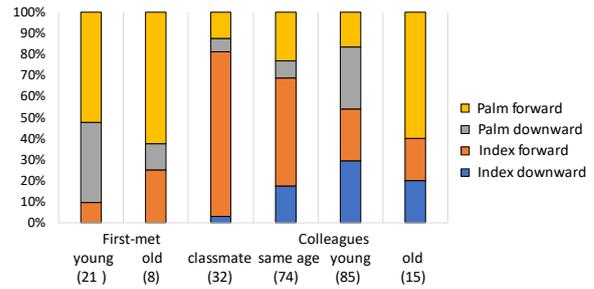


Figure 6. Pointing gestures types (hand shape and motion) and inter-personal relationship

The results in Figs. 5 and 6 show that the index finger is used most frequently towards classmates and colleagues with same age, while open palm is more frequently used towards first-met person (both younger and older), and towards older colleagues. Regarding the motion direction, forward motion is predominant for all inter-personal relationship categories, while downward motion appears with relatively higher frequency towards younger dialogue partners.

Fig. 7 shows the distributions of the pointing gesture functions for different inter-personal relationship categories.

Some trends can be observed in the results of Fig. 7. For example, the “ask” category appears with higher frequency in first-met younger person and older colleagues, i.e., when inter-personal relationship is not very close or very distant. The “joke” category and “others” (including some negative categories like blaming and looking down) appear in interlocutors with close relationship (classmate, same age colleagues and younger colleagues), and in first-met younger person. Regarding the “sympathy” category, it appeared with higher frequency for younger colleagues. However, almost all of these occurrences were by one of the speakers. The results above for attitudes need deeper investigations on larger amount of data.

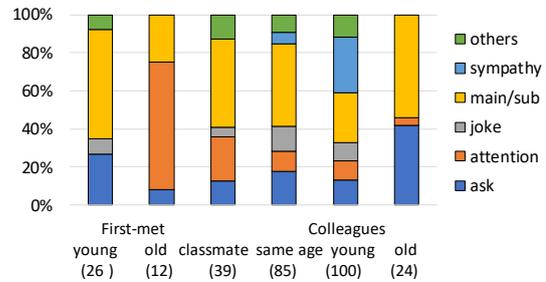


Figure 7. Pointing gestures functions and inter-personal relationship.

Linguistic and functional properties of the “main” and “sub” categories were analyzed in more detail. Fig. 8 shows the distributions for the sub-categories found for the “main” and “sub” categories. Only 9% co-occurred with the name of the interlocutor (“name”), and 5% co-occurred with the pronoun “you”, “he/she”, “your”, “his/her” (“pronoun”). The majority of 75% of the pointing gestures occurred as a replacement of the pronoun, i.e., with the absence of the pronoun or the name of the interlocutor (“pronoun omitted”). The remaining 11% appeared with utterances related to the topics or contents the interlocutor has spoken (“topic”: it means like “I’m talking about what you said”).

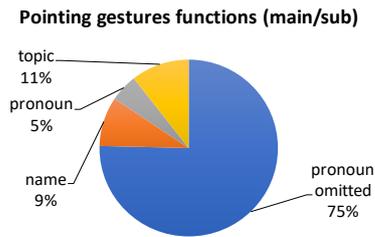


Figure 8. Distributions of functions carried by pointing gestures in the "main" category.

#### 4. Discussion

The analysis results in the previous section indicated that the use of different types of pointing gestures are affected by the inter-personal relationship between the dialogue partners.

Regarding the results for motion direction, the results of high occurrences rates of downward motions may have different reasons for younger colleagues and younger first-met person. In the case of younger colleagues, index downward motions occur with high frequency. Considering that pointing with the index finger in a downward direction from a high position may be perceived as a "bossy" (condescending) behavior, this type of gesture would be more acceptable towards younger interlocutors with close relationship, and less acceptable for others. On the other hand, the downward motions are found with high frequency in first-met younger interlocutors, but with the open palm. For open palm faced up, downward motions are expected to be perceived as a more polite behavior in comparison to the forward counterpart.

Another observation is that the pointing gestures towards older dialogue partners is less frequent than those for younger ones. This is thought to be due to the fact that a pointing gesture might be perceived as impolite towards an older person, while the reverse is thought to be more allowable.

Some of the speakers clearly distinguished the use of the open palm and the index finger within the same dialogue session depending on the inter-personal relationship with the interlocutor (first-met person or older/younger colleague). On the other hand, one of the speakers used open palm with more frequency for all speakers regardless the inter-personal relationship. Thus, the preference of the pointing gesture type is a factor related to the speaker individuality.

Overall, the predominance of the use of index finger instead of the open palm is thought to be because of the predominance of dialogues among people with close relationship and casual conversations. One can expect that the use of open palm will be more frequent in less casual dialogue situations. We plan to investigate more closely on this in future work.

Another topic for future investigation is about generalization of these results for other languages and cultures. In the case of Japanese, the subject "you" is often omitted, so that the pointing gesture usually co-occur with the utterances related or directed to the interlocutor. It would be interesting to check in other languages how often the pointing gestures co-occur with the subject "you".

#### 5. Conclusion

In this study, we analyzed person-directed pointing gestures appearing in a three-party spontaneous dialogue database.

Relationship was found between the gesture type (index finger or open palm, hand orientation, motion direction), and contextual information such as inter-personal relationship and the attitudes conveyed by the pointing gesture.

More than half of the observed pointing gestures were realized with the index finger towards the interlocutor, without being particularly be perceived as impolite or rude behaviors. Regarding the pointing gesture type and the inter-personal relationship, pointing with the index finger moving in the forward direction was found to be predominant towards dialogue partners with close relationship (classmates and colleagues with same age), while pointing with the open palm was found to be more frequent towards first-met person or older person.

Regarding the functions and attitudes conveyed by the pointing gestures, most of them occurred along with utterances whose contents are related or directed to the pointed person, with the subject omitted (i.e. without explicitly uttering a pronoun or name of the pointed person). About 60% of the pointing gestures were accompanied with attitudinal expressions such as yielding the turn (at the end of yes-no questions), sympathizing (at agreeable backchannel utterances), attention drawing (at the beginning of utterances) and joking/bantering.

In future work, we plan to apply the analysis results in the present study to human-robot interaction, so that different pointing gesture types can be appropriately selected according to the inter-personal relationship.

#### 6. Acknowledgements

This study was partly supported by JST, ERATO, Grant Number JPMJER1401 and JST-Mirai Program Grant Number JPMJMI18C6. We thank Taeko Murase, Yuka Nakayama, and Kyoko Nakanishi for their contributions to speech and motion data analyses.

#### 7. References

- [1] C. Ishi, H. Ishiguro, N. Hagita (2013). Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication* 57, No.2014, 233–243, June 2013.
- [2] C. Ishi, H. Hatano, H. Ishiguro (2016). "Audiovisual analysis of relations between laughter types and laughter motions," Proc. of the 8th international conference on Speech Prosody (Speech Prosody 2016), pp. 806-810, May, 2016.
- [3] C. Ishi, T. Minato, H. Ishiguro (2017). "Motion analysis in vocalized surprise expressions," Proc. *Interspeech 2017*, pp. 874-878, Aug. 2017.
- [4] C.T. Ishi, R. Mikata, H. Ishiguro (2018). "Analysis of relations between hand gestures and dialogue act categories," Proc. 9th International Conference on Speech Prosody, 473-477, June 2018.
- [5] C. Liu, C. Ishi, H. Ishiguro, N. Hagita (2013). Generation of nodding, head tilting and gazing for human-robot speech interaction. *International Journal of Humanoid Robotics* (IJHR), vol. 10, no. 1, January 2013.
- [6] C. Ishi, T. Funayama, T. Minato, H. Ishiguro (2016). "Motion generation in android robots during laughing speech," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2016)*, pp. 3327-3332, Oct., 2016.
- [7] C. Ishi, T. Minato, H. Ishiguro (2017). "Motion analysis in vocalized surprise expressions and motion generation in android robots," *IEEE Robotics and Automation Letters*, Vol.2, No.3, 1748 - 1754, July 2017.
- [8] C.T. Ishi, D. Machiyashiki, R. Mikata, H. Ishiguro. (2018) A speech-driven hand gesture generation method and evaluation in android robots. *IEEE Robotics and Automation Letters*, Vol.3, No.4, 3757 - 3764, July 2018.

- [9] A. Kendon (1980) “Gesticulation and speech: two aspects of the process of utterance.” In M. R. Key (ed), *The Relationship of Verbal and Nonverbal Communication*, The Hague: Mouton and Co, pp.207-227, 1980.
- [10] D. McNeill, (1992). *Hand and Mind: What Gestures Reveal about Thought*, Chicago and London: The University of Chicago Press, 1992.
- [11] J. Cassell, M. Stone, and H. Yan. (2000) “Coordination and context-dependence in the generation of embodied conversation,” In Proc. of the *First International Conference on Natural Language Generation*, 2000.
- [12] T. Sowa and I. Wachsmuth (2005). “A model for the representation and processing of shape in coverbal iconic gestures.” In Proc. *KogWis05*, pages 183–188, 2005.
- [13] Bergmann, K., and Kopp, S. (2009) “GNetlc - Using Bayesian Decision Networks for Iconic Gesture Generation” in Proc. of the *9th International Conference on Intelligent Virtual Agents*, Ruttkay, Z., Kipp, M., Nijholt, A., and Vilhjálmsson, H. eds. LNAI, vol. 5773, (Berlin/Heidelberg, Germany: Springer), pp.76-89, 2009.
- [14] Y.I. Nakano, M. Okamoto, D. Kawahara, Q. Li, T. Nishida. (2004). “Converting Text into Agent Animations: Assigning Gestures to Text,” In Proc. *Human Language Technology Conference of the North American Association for Computational Linguistics (HLT-NAACL 2004)*, pp. 153–156.
- [15] Loehr, D. (2004). *Gesture and Intonation*. Washington DC: Georgetown University, PhD Dissertation.
- [16] R. Fang, M. Doering, and J. Y. Chai (2015) “Embodied collaborative referring expression generation in situated human-robot interaction,” in *10th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2015)*, 2015, pp. 271–278.
- [17] A. Sauppé and B. Mutlu, (2014) “Robot deictics: How gesture and context shape referential communication,” in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 2014, pp. 342–349.