Proceedings of ONION 2020: Workshop on peOple in laNguage, vIsiOn and the miNd, pages 7–14
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Analysis of Body Behaviours in Human-Human and Human-Robot Interactions

Taiga Mori[*†], Kristiina Jokinen[†], Yasuharu Den[‡]
[*]Graduate School of Humanities and Studies on Public Affairs, Chiba University
[‡]Graduate School of Humanities, Chiba University
1-33 Yayoicho, Inage-ku, Chiba 263-8522, Japan

[†]AI Research Center, AIST Tokyo Waterfront
2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

mori-taiga@aist.go.jp, kristiina.jokinen@aist.go.jp, den@chiba-u.jp

Abstract

We conducted preliminary comparison of human-robot (HR) interaction with human-human (HH) interaction conducted in English and in Japanese. As the result, body gestures increased in HR, while hand and head gestures decreased in HR. Concerning hand gesture, they were composed of more diverse and complex forms, trajectories and functions in HH than in HR. Moreover, English speakers produced 6 times more hand gestures than Japanese speakers in HH. Regarding head gesture, even though there was no difference in the frequency of head gestures between English speakers and Japanese speakers in HH, Japanese speakers produced slightly more nodding during the robot's speaking than English speakers in HR. Furthermore, positions of nod were different depending on the language. Concerning body gesture, participants produced body gestures mostly to regulate appropriate distance with the robot in HR. Additionally, English speakers produced slightly more body gestures than Japanese speakers.

Keywords: human-human and human-robot interactions, hand gestures, head gestures, body gestures, Japanese and English

1. Introduction

In recent years, multimodal interaction has become a much studied research area and many investigations have been conducted to widen our understanding of human behaviour and interaction dynamics. Research concerns multimodal resources and models on various aspects of interaction associated with the use of whole body and the combination of visual and auditive modalities, and recently also novel technology has offered interesting possibilities for analysing human behaviour in an accurate manner: the use of video, motion capture, eye-tracker, and many sensor devices provide data which can be used as input to bigdata and machine-learning calculations in order to establish accurate correlations and relations among the modalities. Moreover, novel applications such as interactive social robots have also become common, and in order to develop more natural systems that can understand human behaviour as well as produce expressive and engaging behaviour, it is important to study multimodal communication in situations with humans and other interactive agents. For instance, co-speech gesturing is important in making one's presentation natural, engaging, and expressive, and it is also important to be able to detect and interpret the relevant signals so as to understand the partner's communicative intentions.

In this paper, we focus on gesturing to study spoken interactions in a practical context of instructing or giving advice to a colleague, about how to perform a particular care-giving task. In our research we have selected hand gestures and head nodding as the primary object of study. There is already much research on how gestures and nods function in human communication, while coordination of speech and gestures is less studied, especially for the purpose of human-robot interaction. Important goals of our research are thus related to deepening our knowledge of the use of co-speech gestures in interaction, and to investigate how to build models for enabling more natural interaction with robots. Such multimodal interaction models can be applied in human-robot interaction. We annotated the gestures using a modified MUMIN annotation scheme (Allwood et al. 2007). The scheme uses gesture features divided into form and function features, and it is described more in Section 3 and Section 8. The research question concerns how to use gesturing in grounding information and creating mutual understanding of the discussion topic, i.e. how the user's gestures can be used to establish an appropriate way to continue the interaction. We will especially study differences between human-human and human-robot interaction and also compare interactions conducted in English and in Japanese. Our hypotheses with respect to gesturing are:

1) There are more body movements in HH than in HR dialogues.
2) In particular, there are more hand gestures in HH than in HR dialogues, and there are more body movements in HR than in HH.
3) There are more body movements in dialogues conducted in English than in Japanese.
4) There are more body movements when speaking than in listening.
5) There is correlation between body movements and the person's perception of the dialogue in general.

We will also combine presentation of spoken information with gesture and (later) eye-gaze information to design the system's behaviour with respect to multimodal information. For instance, in the robot's listening side suitable dialogue strategies are available to predict the user's understanding or misunderstanding based on their gesture reaction and to specify the presented information appropriately. On the generation side, dialogue strategies include multimodal signals to provide a relevant response and present information and mark the speaker's continued attention to the partner. This kind of grounding in interaction (Clark and Schaefer 1987) is important in understanding the partner's intentions and making one's own intentions known, i.e. to enable smooth interaction. It is hypothesized that the robot's perceived cooperation and grounding of information improves naturalness of its spoken interaction. This is crucial especially in long-term interaction (Heylen et al. 2010) and in various applications related to social robotics where the robot is to act like a co-worked or companion and provide information to the user as well support natural, friendly interaction: the robot's detection of the user's understanding and misunderstanding is important to provide expressive interaction which supports emotionally satisfying and pleasant interaction (Kanda et al. 2004; Beck et al 2010). We were interested in the user's and the robot's mutual understanding process, and especially how the non-expected and misunderstood situations are reflected in the user's gesture patterns, to be able to use this information in designing the robot's interaction strategy.

This paper is structured as follows. First, a short overview of relevant gesture studies is reviewed in Section 2, then the data and annotation scheme are presented briefly in Section 3, and preliminary results shown in Section 4. Next, some methodological issues as well as ethical issues related to the monitoring and data collection in the context of interactive systems are discussed in Section 5, and conclusions are drawn in Section 6. Finally, specific annotation scheme is attached to Section 8.

2. Overview of Previous Work

In linguistic interaction, speech is commonly associated with gesturing (Kendon 2004) and co-speech gestures have been studied from the point of view of turn-taking (Duncan 1972; Streek 2009), iconic gestures and description (Lis and Navarretta 2014), pointing gestures (Jokinen 2010), gestures and multimodal information (Paggio and Navarretta 2013), gestures and neurocognitive processes (Kita et al. 2017), and intercultural comparison (Navarretta et al. 2012; Endrass et al. 2011). Also, in integrating more natural interaction possibilities for a robot (Jokinen and Wilcock 2014; Ono et al. 2001). Automatic analysis platforms have also been developed (Heimerl et al. 2019) and machine learning is used to study interpersonal dynamics (Baltrušaitis et al. 2019). In human-robot interaction (HRI), multimodal issues are also important as speaking robots start to appear in homes, public spaces, and work. The robot's communicative patterns are still rather inflexible, and user evaluations usually point to the robot's inflexible feedback strategies and monotonous engagement with the human. Social robots range from speaking heads (Alexa, Google) to more dialogue-oriented interactive systems for task-based scenarios (Sidner et al. 2015; Jokinen et al., 2018) and although much research is conducted on speech-based HRI, low utilization of multimodal signal in HRI still constrains the understanding of the role of social signals in HR.

3. Data and Annotation

The data is from the AICO Corpus (Jokinen, 2020) which is available for cooperative research at AIST. It consists of 30 participants, 20 native Japanese and 10 English speakers with backgrounds in Europe, US and South-East Asia, of which 10 are women. They are students and researchers, aged 20-60, and they have experience on IT but no experience on robots. Figure 1 shows the experimental setup. Each participant had two sessions one with a human partner and one with a robot partner for about 10 minutes respectively, so altogether there are 60 interactions, i.e. 30 human-human (HH) and 30 human-robot (HR) interactions. In HH session, one of the experimenters played the role of the human partner and the Nao robot played the role of the robot partner in HR session. Other experimenters monitored the session from the next room to intervene when problems arise. Data was collected using video camera, Kinect, eye-tracker and a questionnaire about impression on the robot. The setup is described in more detail in Ijuin et al. (2019) and Jokinen (2019). This data enables us to compare interaction patterns across the human and agent partners. In this paper we compare the human-human and human-robot interactions, and also draw some observations concerning interactions conducted in Japanese and in English.

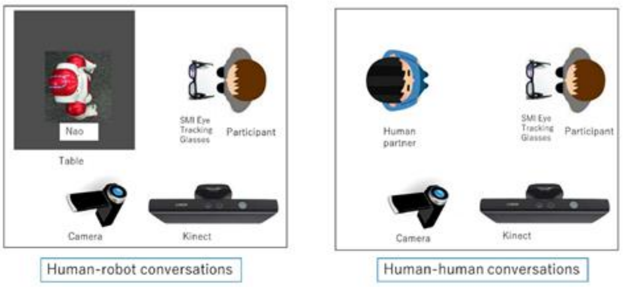

Figure 1: The experimental setup

Gestures can be classified according to a modified version of the MUMIN annotation scheme (Allwood et al. 2007), which is based on the gesture form (e.g., up-open, curled-fingers and extended-finger) and the gesture function (e.g., iconic, deictic and emblem gestures). For the full set of annotation categories, see Section 8. At the moment, 19 interactions have been annotated and 16 of them were used for the following analyses. Table 1 shows the breakdown of the analysed data.

Japanese				English			
HH		HR		HH		HR	
M	F	M	F	M	F	M	F
2	1	4	1	2	2	2	2

Table 1: Breakdown of analysed data

M and F mean male and female participant's number respectively.

4. Gesture and Body Posture Analysis

4.1 Hand Gestures

4.1.1 Mean Frequency of Hand Gestures

Figure 2 shows the mean frequency of hand gestures. As can be seen, hand gestures considerably decreased in HR, which is in accordance with our hypothesis. Considering the language differences, it is interesting that even though the English speakers produced 6 times more hand gestures in HH than the Japanese, their difference is not so big in HR. The similar trends between English and Japanese in HR is because both English and Japanese speakers produced only self-directed gestures in HR, such as touching a table or scratching one's body. This implies that for realizing natural interaction with a robot, it is necessary to focus first on eliciting gestures from the user rather than on recognizing gestures.

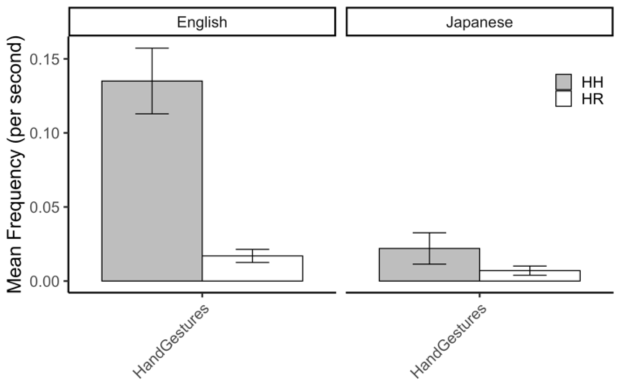

Figure 2: Mean frequency of hand gestures. Error bars show the standard errors.

4.1.2 Form

The most frequent hand gesture form for the English speakers was *curled-fingers* in both HH and HR sessions (Figure 3). However, in HH, other forms also occurred, while this form was almost the only one observed in HR. In our scheme, *curled-fingers* is defined as the default form realized without any effort, in contrast to the *opening a palm* or *pointing* (Table 2). That is, English speakers made more complex hand forms in HH. Similar pattern was also observed for the Japanese speakers, although they produced less hand gestures than English speakers. Considering the language differences, English speakers produced twice more gestures than Japanese in almost all hand form. On the basis of this result, it can be said that English interaction is more dependent on hand gestures than Japanese interaction. This suggests that robots have to recognize more various hand gesture forms in English than in Japanese.

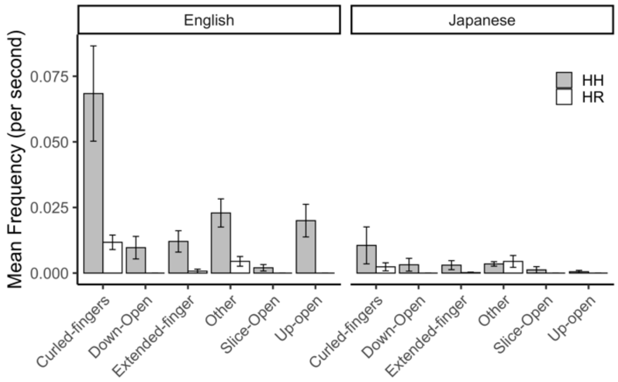

Figure 3: Mean frequency of hand gesture forms

4.1.3 Function

As with the gesture forms, the functions of hand gestures were also more diverse in HH than in HR (Figure 4). Almost all gestures that were produced in HR are classified as *adapter* gestures, such as leaning one's body weight onto a table or touching one's body. As *rhythmic*, *iconic*, *deictic* and *emphasis* gestures are obviously more interactive than *adapter* gestures, it can be concluded that the participants mostly produced other-directed gestures in HH. Considering the language differences, English speakers produced more rhythmic gestures than Japanese, suggesting that prosodic information including intonation and rhythm might be more important in English than in Japanese. Another important implication is that Japanese speakers might emphasise important point in other modality because they produced fewer *emphasis* hand gesture. In conclusion, because Japanese speakers produce less other-directed hand gestures than English speakers, the need for robots to accurately recognize hand gesture might be lower in Japanese than in English.

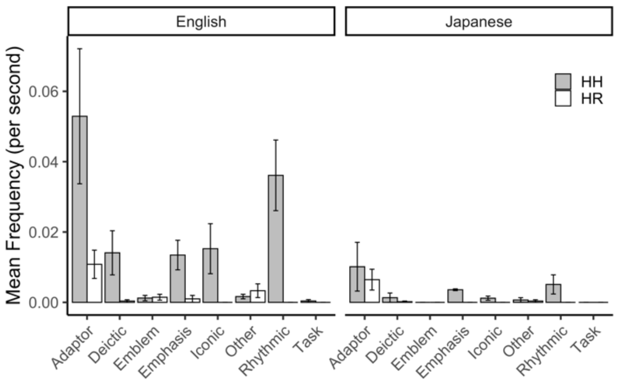

Figure 4: Mean frequency of hand gesture functions

4.1.4 Trajectory

Concerning the trajectory of the gestures, the *straight* trajectory is the most frequent in HH interactions: *complex* trajectories occurred only about half as many as the *straight* ones (Figure 5). However, in HR interactions, *complex* trajectory is not observed at all. This observation is consistent with the fact that there are very few complex gesture forms in HR. *Complex* gesture trajectories and forms could represent more rich information visually, but they would demand more cognitive costs in terms of production, recognition and interpretation. In the case of HR interaction, the participants seem to "save" the cost of producing complex gestures, because they did not regard the robot as a partner who can recognize rich visual information. In order to elicit hand gestures from humans in HR interaction, the robot should produce gestures naturally so that the human partners can assume and perceive that it can interact with them tactfully in visual modality.

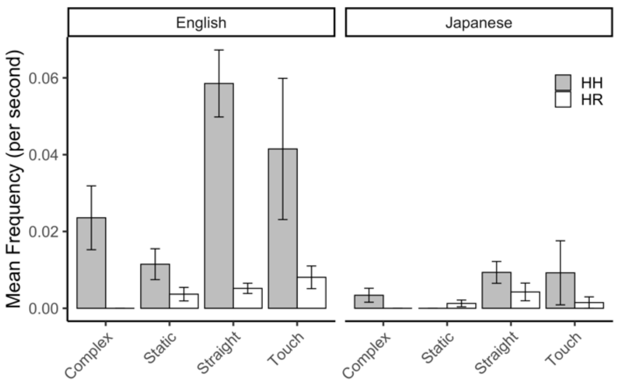

Figure 5: Mean frequency of hand gesture trajectories

4.1.5 Handedness

Based on the results concerning the hand gesture form and trajectory, it can be predicted that there would be less gestures using both hands than gestures using a single hand, because both hand gestures would be more costly. However, contrary to the prediction, the difference between single and both hands gesturing was not so big either in HH nor in HR (Figure 6). This suggests that both hand gestures cannot be omitted into single hand gestures because they are determined by their function and the content of the gesture expression. For instance, one participant represented 'low' and 'high' with the left hand and the right hand respectively; this gesture could not be represented only with a single hand.

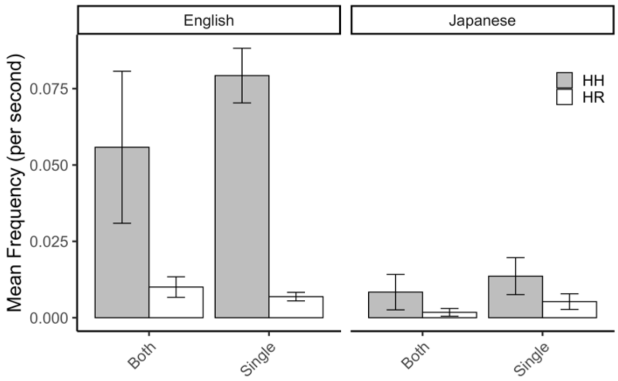

Figure 6: Mean frequency of handedness

4.1.6 Repetition

Similarly to the handedness, there was no difference between *single* and *repeated* gestures in either sessions (Figure 7). *Single* gestures were frequently observed in *emphasis* gestures, while *repeated* gestures were frequently observed in *rhythmic* gestures.

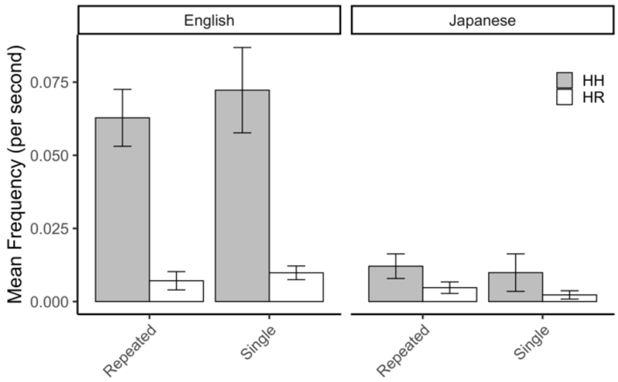

Figure 7: Mean frequency of hand gesture repetition

4.2 Head Gestures

4.2.1 Mean Frequency of Head Gestures

Figure 8 shows mean frequency of head gestures. As can be seen, head gestures decreased in HR in a similar manner as hand gestures. While there was not so big difference between English and Japanese in HH, Japanese speakers produced slightly more head gestures in HR than English speakers.

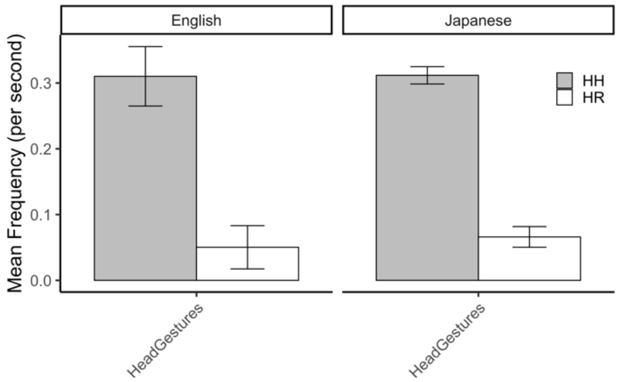

Figure 8: Mean frequency of head gesture

4.2.2 Form

Nod gestures were remarkably most frequent in HH (Figure 9). Although Maynard (1989) showed that native Japanese speakers tend to nod more frequently than American English speakers, there was no big difference between Japanese and English in HH. The following point can be given as reasons for this. Because not all English participants were native speakers, their interactional manner in their first language produced this incoherent result. As evidence for this, individual differences were larger in English speaking interactions than in Japanese interactions. On the other hand, nod gestures observed in HR were slightly more frequent in Japanese. One possible reason is that the Japanese speakers behaved with the robot in the same way as they always do with the human partner. Moreover, Japanese nodded with a response token overlapping partner's utterance while English speakers nodded silently. It is interesting to analyse if the Japanese nodded at the same position in the partner's utterance in HHI and HRI, and also to analyse the relationship between response token and head gestures.

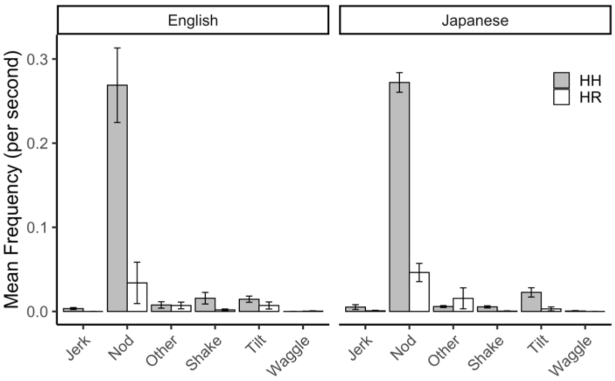

Figure 9: Mean frequency of head gesture forms

4.2.3 Function

Regardless of language, *acknowledge* gestures were most frequent in HH, and *emphasis* gestures were the second most frequent gestures (Figure 10). On the other hand, in HR, the *acknowledge* and *adapter* gestures were relatively more frequent than other functions. Almost all *acknowledge* gestures were observed as *nod*. In HR, Japanese speakers produced more *acknowledge* gestures than English, due to the fact that the Japanese did not nod only during human speaking but also when the robot speaking. Japanese speakers also produced more nods towards the end of their utterance than the English speakers. This implies that Japanese monitored the partner more strictly to elicit gestures when they had the turn. That is to say, robots have to recognize and response to that.

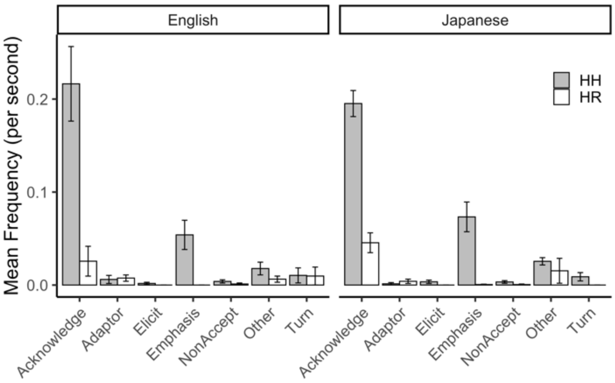

Figure 10: Mean frequency of head gesture functions

4.2.4 Repetition

While *repeated* gestures were more frequent than *single* ones in HH, this tendency was reversed in HR (Figure 11). Although *repeated* gestures would involve more physical cost than *single* ones, they also enable us to represent strong empathy or deep understanding to a speaker. Participants intended to give strong encouraging feedback to the partner in HH, but they saved the cost when talking to the robot. Moreover, the fact that this tendency is common to English and Japanese speakers implies that the function of repetition is common to English and Japanese. In other words, robots can interpret the repetition of head gestures in the same way between Japanese and English.

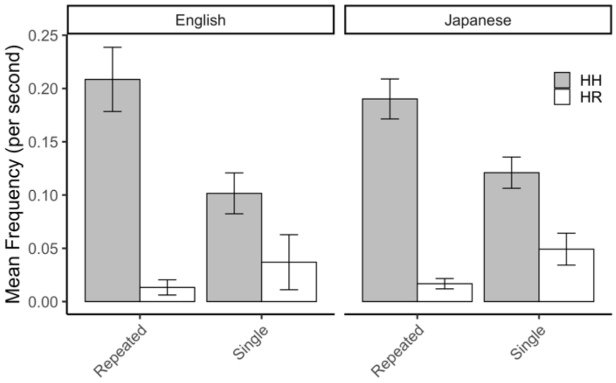

Figure 11: Mean frequency of head gesture repetition

4.3 Body Gestures

4.3.1 Mean Frequency of Body Gestures

Figure 12 shows mean frequency of body gestures. While hand and head gestures were more frequent in HH, body gestures were more frequent in HR. This result suggests that it is more necessary for robots to recognize body gestures than hand and head gestures. Even though there was not so big difference, English speakers produced more body gestures than Japanese in accord with hypothesis.

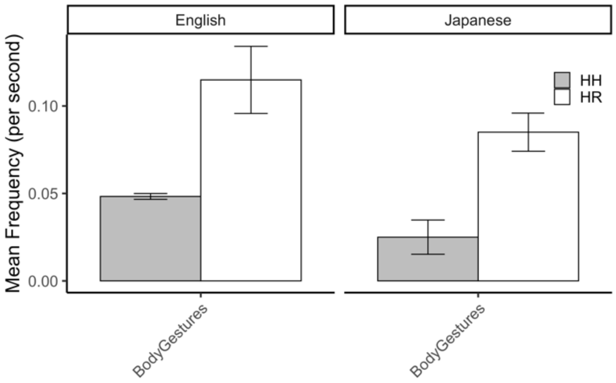

Figure 12: Mean frequency of body gestures

4.3.2 Form

Forward movements were observed most frequently in HR (Figure 13). For instance, participants leaned toward the robot when they spoke to the robot. They sometimes spoke to the robot in the middle of its utterance even though it was programmed to light up and sound on the end and start of its turn. This implies that they could not use unnatural cues for turn-taking. On the other hand, *backward* movements were observed, for instance when participants leaned backward because the robot failed to catch their words or behaved unexpectedly, and then returned to the original position in order to restart the conversation. These behaviours may imply that they made interactive formation with the robot like an F-formation (Kendon 2004), i.e. they broke away from the interactive situation when the robot failed to behave as expected. As for other movements, such as moving sideways and changing body weight from one foot to another were observed in both HH and HR, but shaking one's legs angrily was observed in only HR.

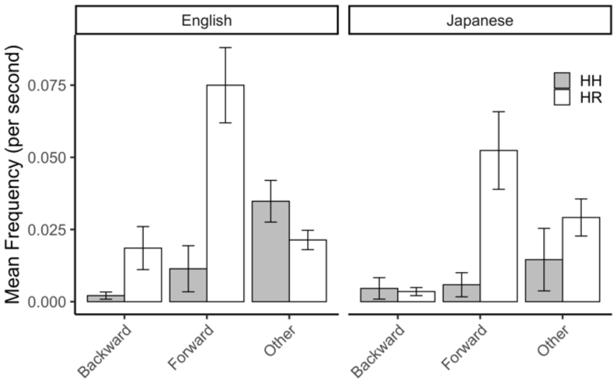

Figure 13: Mean frequency of body gesture forms

4.3.3 Function

The most frequent body function was *better contact* in HR regardless of language (Figure 14). Participants were able to regulate appropriate distance each other in HH, however they had to do that by oneself in HR, and which involves cost for humans. It is desirable for robots in the future to recognize and regulate appropriate distance to humans oneself. Moreover, although frequency of *adapter* gestures was equal in HH and in HR, the gestures occurred in different occasions. While participants frequently changed their body posture when nervous in HH, they shook their legs in frustration to the robot's failure in HR. The data, although small to draw generalisations, shows that male participants looked irritated and produced more *adapter* gestures when the robot failed to catch their words, while females just behaved as confused or laughed. Based on this, it can be assumed that females perceived the robot as more "social entity" than males.

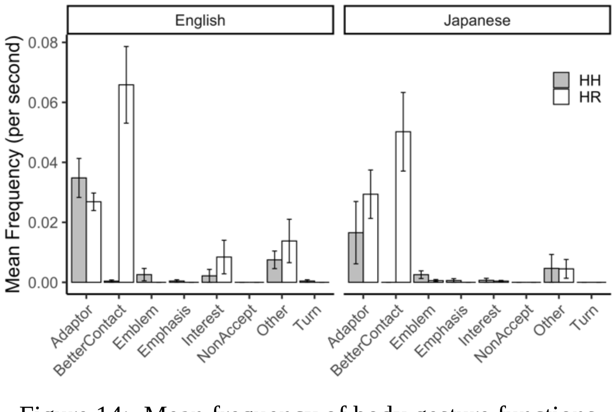

Figure 14: Mean frequency of body gesture functions

4.3.4 Repetition

Single gestures were most frequent in HH and in HR (Figure 15). They were observed when participants leaned forward with each utterance to speak to the robot, or they changed their body weight from one foot to another. *Repeated* gestures were observed as swaying body co-occurred with *rhythmic* hand gesture, or shaking legs from stress. *Static* gestures were observed when participants continued a head forward posture for a few second to reduce physical costs of leaning forward repeatedly. It also suggests that it is larger cost to regulate physical distance. Consequently, body gestures might have involved transmission of information rather than content of interaction, and reflected their mental state such as being nervous or frustrating because they have less expressiveness than hand and head gestures.

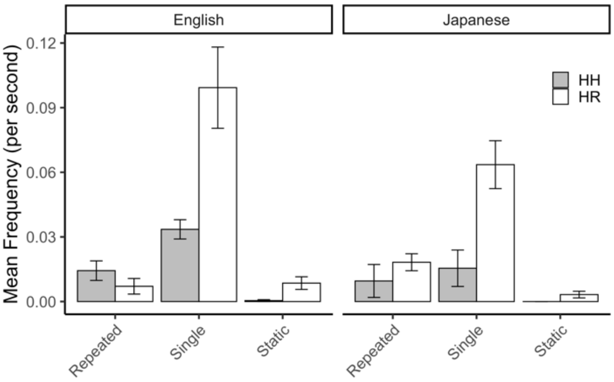

Figure 15: Mean frequency of body gesture repetition

5. Future Work

Since the annotated data is fairly small, we first aim to finish the annotations so to provide a solid basis for the statistical analysis. In the next step of the research, we plan to study time correlation between speech and multimodal gesturing (co-speech hand gesturing, nodding, and body movements). We will focus on the use of visual and auditory information in order to build a model for anticipating the partner's gestures and their timing within the spoken interaction, possibly combined with a functional meaning of the gesture. Our goal is to investigate how auditive and visual modalities are used as communicative signals in various interactive situations, and how to learn interaction models which can ultimately be applied to develop natural human-robot interaction (cf. Beck et al., 2010, Jokinen et al. 2014).We are especially interested in time correlation between response token and head gestures, because it is known that recipient's nod usually co-occurs with response token in Japanese. A lot of previous studies attempted to predict some features of response token from precedent utterance to develop voice interactive system in Japanese. However, it is necessary to reveal, for instance the relationship between prosodic features of response token and the depth of nod, or the location of nod on the co-occurred response token in order to develop multimodal interactive system.

Finally, we plan for a comparison of the results using different corpora. It will be useful to compare various interactive situations and extract features that enable us to generalise over relevant attributes in interactive situations and also to explore methodological issues related to modelling and processing human physical characteristics.

6. Acknowledgement

We would like to thank all the participants who took part in the experiments as well as to our colleagues for their assistance in the preparation, collection, and post-processing of the data. This paper is based on results obtained from Future AI and Robot Technology Research and Development Project commissioned by the Japan New Energy and Industrial Technology Development Organization (NEDO).

7. References

Allwood, J., Cerrato, L., Jokinen, K., Navarretta C., Paggio,P.: The MUMIN Coding Scheme for the Annotation of Feedback, Turn Management and Sequencing Phenomena. Multimodal Corpora for Modelling Human Multimodal Behaviour. Special issue of the International Journal of Language Resources and Evaluation, 41(3-4), 273-287(2007). SpringerLink Online: http://www.springerlink.com/content/x745801041m52553/fulltext.pdf

Baltrušaitis, T., Ahuja C., Morency, L.: Multimodal Machine Learning: A Survey and Taxonomy, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423-443, 1 Feb. 2019.

Beck, A., Canamero, L., Bard, K.A.: Towards an Affect Space for Robots to Display Emotional Body Language, in Proceedings of the 19th IEEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN 2010), Principe di Piemonto -Viareggio, Italy, 2010.

Clark, H. H., Schaefer, E. F.: Collaborating on contributions to conversation. Language and Cognitive Processes, 2, 19-41 (1987).

Duncan, Jr., S.: Some signals and rules for taking speaking turns in conversations. Journal of Personality and Social Psychology, 23 (2), 283-292 (1972).

Endrass, B., Nakano, Y., Lipi, A. A., Rehm, M., André, E.: Culture-Related Topic Selection in Small Talk Conversations across Germany and Japan. Lecture Notes in Computer Science, 6895, 1-13 (2011).Feldman, R. S., Rim, B.: Fundamentals of Nonverbal Behavior. Cambridge: Cambridge University Press (1991).

Heimerl, A., Baur, T., Lingenfelser, F., Wagner, J., André, E.: NOVA - A tool for eXplainable Cooperative Machine Learning, in Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge (2019).

Heylen, D., Krenn, B., Payr, S.: Companions, Virtual Butlers, Assistive Robots: Empirical and Theoretical Insights for Building Long-Term Social Relationships. In: Trappl, R. (ed.): Cybernetics and Systems 2010, pp. 539–570. Austrian Society for Cybernetic Studies. Vienna, Austria (2010).

Ijuin, K., Jokinen, K., Kato, T., Yamamoto, S.: Eye-gaze in social robot interactions – Grounding of information and eye-gaze patterns. JSAI 2019 (2019)

Jokinen, K.: Constructive Dialogue Modelling – Speech Interaction with Rational Agents. John Wiley & Sons, Chichester, UK (2009).

Jokinen, K.: Dialogue models for Social Robots. In: Proceedings of ICSR'2018, Qingdao, China (2018).

Jokinen, K.: The AICO corpus. Technical Report. AI Research Center, AIST. (2019).

Jokinen, K.: Pointing Gestures and Synchronous Communication Management. In: A. Esposito, N. Campbell, C. Vogel, A. Hussain, A. Nijholt, Eds., Development of Multimodal Interfaces: Active Listening and Synchrony, pp. 33-49. Berlin: Springer (2010)

Jokinen, K., Nishimura, S., Watanabe, K., Nishimura, T.: Human-Robot Dialogues for Explaining Activities. In: Proceedings of IWSDS-2018, Singapore (2018).

Jokinen, K., Wilcock, G.: Multimodal Open-Domain Conversations with the Nao Robot. In: Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialogue Systems into Practice Springer, New York, pp. 213–224 (2014).

Kanda, T., Hirano, T., Eaton, D., Ishiguro, H.: Interactive robots as social partners and peer tutors for children: A field trial. Human-Computer Interaction, 19 (1), 61-84 (2004).

Kendon, A.: Gestures: Visible Action as Utterance. Cambridge: Cambridge University Press (2004).

Kita, S., Alibali, M. W., Chu, M.: How Do Gestures Influence Thinking and Speaking? The Gesture-for-Conceptualization Hypothesis. Psychological Review, 124(3), 245-266. (2017).

Lis, M., Navarretta C.: Classifying the form of iconic hand gestures from the linguistic categorization of co-occurring verbs. In: Proceedings of the European Symposium on Multimodal Communication (MMSym'13), Valetta, Malta (2014).

Maynard, S.: Japanese conversation: Self-contextualization through structure and interactional management. Norwood, NJ: Ablex (1989).

Navarretta, C., Ahlsén, E., Allwood, J., Jokinen, K., Paggio, P.: Feedback in Nordic First-Encounters: a Comparative Study. LREC 2012: 2494-2499

Navarretta, C., Ahlsen, E., Allwood, J., Jokinen, K., Paggio, P.: Feedback in Nordic firstencounters: a comparative study. Proceedings of 8th International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, 2012.

Paggio, P., Navarretta, C.: Head movements, facial expressions and feedback in conversations: empirical evidence from Danish multimodal data. J. Multimodal User Interfaces 7(1-2): 29-37 (2013)

Ono, T., Imai, M., & Ishiguro, H. (2001). A Model of Embodied Communications with Gestures between Human and Robots. Proceedings of the Annual Meeting of the Cognitive Science Society, 23.

Senft, E., Baxter, P., Kennedy, J., Lemaignan, S., Belpaeme, T.: Supervised autonomy for online learning in human-robot interaction. Pattern Recognition Letters, 99: 77-86 (2017).

Sidner, C., Rich, C., Shayganfar, M., Bickmore, T., Ring, L. and Zhang, Z.: A Robotic Companion for Social Support of Isolated Older Adults. Procs of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction, 289-289 (2015).

Streeck, J.: Gesturecraft: The Manufacture of Meaning. Amsterdam: John Benjamins (2009).

Wilcock, G., Jokinen, K.: Advances in Wikipedia-based Interaction with Robots. ICMI Workshop Multi-modal, Multi-Party, Real-World Human-Robot Interaction, pp.13-18 (2014).

8. Appendix: AICO Annotation Scheme

8.1 Hand Gestures

8.1.1 Form

Hand gesture forms are classified into following 6 types based on the shape of a palm or fingers.

Up-open	Opening a palm upwards
Down-open	Opening a palm downwards
Sliced-open	Opening a palm sideways
Extended-finger	Extending a finger towards pointing
Curled-fingers	Curling fingers close to palm
Other	Gesture form not listed above

Table 2: Classification of hand gesture forms

8.1.2 Function

Hand gesture functions are classified into following 8 types in terms of communicative function.

Deictic	Pointing to a concrete or an abstract referent
Rhythmic	Giving rhythm to speech
Emphasis	Emphasising a particular point in talk
Iconic	Describing concrete or abstract objects
Emblem	Expressing a particular symbolic meaning that is culturally conditioned
Task	Performing a task
Adapter	Improving comfort or reducing stress
Other	Gesture function not listed above

Table 3: Classification of hand gesture functions

8.1.3 Trajectory

Hand gesture trajectory means the movement path of the gesturing hand. Those trajectries are classified into following 4 types.

Straight	Moving up, down or sideways
Complex	Complex directions
Static	Staying in the same position and location
Touch	Like static but keep touching

Table 4: Classification of hand gesture trajectories

8.1.4 Handedness

Handedness is decided based on whether the gesture is performed with one or both hands.

Both	Both hands
Single	Single hand

Table 5: Classification of handedness

8.1.5 Repetition

Hand gesture repetition means whether the gesture is composed of a single movement or several similar movements.

Single	Single movement
Repeated	Repeated movements

Table 6: Classification of hand gesture repetition

8.2 Head Gestures

8.2.1 Form

Head gesture forms are cassified into following 6 types based on the movements of the gesturing head.

Jerk	Moving sudden up
Nod	Moving up-down
Shake	Rotating side-to-side
Tilt	Tilting on one side
Waggle	Moving sideways
Other	Gesture form not listed above

Table 7: Classification of head gesture forms

8.2.2 Function

Head gesture functions are classified into following 7 types in terms of communicative function.

Acknowledge	Giving encouraging feedback to the partner
NonAccept	Objecting or withdrawing from what the partner is saying or doing
Emphasis	Emphasising some particular point in talk
Turn	Giving turn to the partner or accepting turn from the partner
Adapter	improving comfort or reducing stress
Elicit	Eliciting feedback from the partner
Other	Head function not listed above

Table 8: Classification of hand gesture functions

8.2.3 Repetition

Head gesture repetition means whether the gesture is composed of a single movement or several similar movements.

Single	Single movement
Repeated	Repeated movements

Table 9: Classification of hand gesture repetition

8.3 Body Gestures

8.3.1 Form

Body gesture forms are cassified into following 3 types based on the movements of the gesturing body.

Forward	Leaning towards the partner
Backward	Leaning away from the partner
Other	Gesture form not listed above

Table 10: Classification of body gesture forms

8.3.2 Function

Body gesture functions are classified into following 8 types in terms of communicative function.

Interest	Giving feedback that shows interest to the partner’s talk
BetterContact	Moving closer to hear or speak clearly to the partner
NonAccept	Objecting or withdrawing oneself from what the partner is saying or doing
Emphasis	Emphasising some particular point in talk
Turn	Giving turn to the partner, or accepting turn
Emblem	Expressing a particular symbolic meaning that is culturally conditioned
Adapter	Improving comfort or reducing stress
Other	Body gesture function not listed above

Table 11: Classification of hand gesture functions

8.3.3 Repetition

Body gesture repetition means whether the gesture is composed of a brief single movement, a long single movement or several similar movements.

Single	Single movement
Repeated	Repeated movements
Static	Staying in the same position and location for few second

Table 12: Classification of hand gesture repetition