電気学会論文誌 C（電子・情報・システム部門誌）
IEEJ Transactions on Electronics, Information and Systems
Vol.137 No.12 pp.1625-1632 DOI: 10.1541/ieejeiss.137.1625

Paper

Gender Effects on Lexical Alignment in Human-Robot Interaction

Mitsuhiko Kimoto[*,**a)] Non-member, Takamasa Iio[*,***] Non-member
Masahiro Shiomi[*] Non-member, Ivan Tanev[**] Non-member
Katsunori Shimohara[**] Member, Norihiro Hagita[*] Non-member

(Manuscript received March 1, 2017, revised July 31, 2017)

Humans tend to change their lexical expressions to resemble those used by their interlocutors to achieve smooth conversations. Such phenomena, called "lexical alignments," are affected by gender. Even though lexical alignment is observed not only in human-human interaction but also in human-robot interaction, the gender effects on it in human-robot interaction haven't been investigated yet. Identifying whether gender affects lexical alignment in human-robot interaction would contribute to the design of conversational strategies for interactive robots for more natural interaction. This paper reveals that gender affects lexical alignment in human-robot interaction. We conducted an experiment with twenty participants who interacted with a robot in object reference conversations and referred to an object whose identity was confirmed by a robot. We developed a robotic system that engaged in object reference conversations with two interaction strategies and measured the gender effects on lexical alignment in human-robot interaction. Our experimental results showed that female participants were lexically more aligned with the robot than males; female participants used more references that were useful to uniquely identify objects in environments than males.

Keywords : human-robot interaction, alignment, gender effects, interaction strategy

1. Introduction

In communication, humans tend to repeat lexical expressions that resemble those of their interlocutor[(2)(3)]. This phenomenon is called lexical alignment. Its importance is located in the fact that it is often associated with successful dialogues. Nenkova et al.[(4)] found that alignment in the use of high-frequency words was correlated with task success and turn-taking in dialogues. Lee et al.[(5)] reported that the alignment measures of two prosodic features, pitch and energy, were higher in positive interactions between married couples than in negative interactions. According to Pickering and Garrod[(6)], alignment is a critical element for successful communication.

Alignment has often been researched in terms of gender effects. Some past research works on human-human interaction observed differences in the degree of alignment by gender[(7)-(9)]. Namy et al.[(8)] reported females aligned in relation to word pronunciation more than males. Levitan et al.[(7)] found that alignment in acoustic/prosodic features was most prevalent for female-male conversation pairs.

Alignment is also observed in interactions between a human and artificial media, for example, in spoken dialogue systems[(10)-(12)] and robots[(13)-(15)]. Iio et al.[(14)] found that lexical alignment and the alignment of word choices occur in conversations between humans and a robot. Their experiment's participants were more likely to use the same words as the robot in conversations. In the human-robot interaction field, lexical alignment findings have also used to improve a robot's performance. For example, Kimoto et al.[(16)] proposed a strategy that incorporates lexical alignment findings in the robot's behavior model to improve the recognition performance of objects indicated by a user.

In alignment between humans and artificial media/robots, other works cited gender differences[(13)(15)(17)]. Thomason et al.[(17)] concluded that males aligned more than females for vocal loudness features. Strupka et al.[(15)] reported that even though humans aligned with robot voices in relation to acoustic energy level, gender had no effect on their voice adjustments.

Past research works reported that the degree of alignment is affected by gender. However, since no study has addressed the gender-based differences of lexical alignment in human-robot interaction, we focus on gender effects on lexical alignment in human-robot interaction. Gender effects on lexical alignment in human-robot interaction is worth investigating for the following two reasons. First, to investigate gender effects is important for understanding human activity and many researchers have investigated gender differences on various psychological attributes[(18)(19)]. Hyde[(19)] stated the reason for the importance of the research on gender differences and similarities in his review article as follows: "… stereotypes about psychological gender differences abound, influencing people's behavior, and it is important to evaluate whether they are accurate". In the research fields of alignment, gender differences have also been investigated. However, gender differences on lexical alignment between people and robots have not been well investigated and not been revealed. Interaction between people and robots is the new interaction style compared to the interaction between people, and the gender effects

This paper is an extended version of a previous work of Kimoto et al.[(1)] and contained additional experiment results about gender effects and more detailed discussions.

a) Correspondence to: Mitsuhiko Kimoto. E-mail: kimoto@atr.jp
* Intelligent Robotics and Communication Laboratories, ATR
2-2-2, Hikaridai, Seikacho, Sourakugun, Kyoto 619-0288, Japan
** Graduate School of Science and Engineering, Doshisha University
1-3, Tatara Miyakodani, Kyotanabe, Kyoto 610-0321, Japan
*** Graduate School of Engineering Science, Osaka University
1-3, Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan

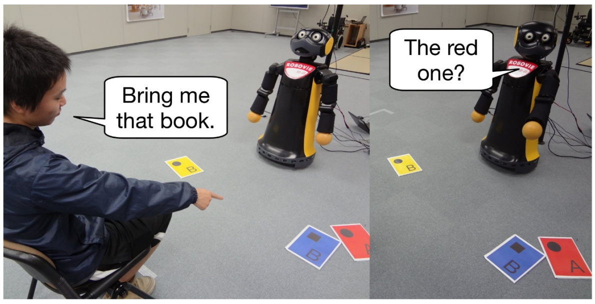

Fig. 1. Recognition of indicated object by an interlocutor

on human-robot interaction would be different from human-human interaction. Revealing gender-based differences of lexical alignment that occur in human-robot interaction has importance itself. Next, recently lexical alignment is used to design human-robot interaction. If we can identify the gender effects on lexical alignment between humans and robots, such understanding would help design human-robot interaction. For example, Kimoto et al.[20] focused on strategies to clarify the lexical expressions of users' references when they refer to the objects in an environment and compared two strategies. In one strategy, robots implicitly elicited the lexical expressions contained in the robot's database from users through lexical alignment (implicit alignment). In their second strategy, robots explicitly instructed a user how to refer to objects (explicit instruction). They concluded that implicit alignment is more effective than explicit instruction. However, they did not compare the gender effects of the two strategies, and if lexical alignment is affected by gender, appropriate strategies may differ based on gender.

This paper investigated whether gender-based differences of lexical alignment occur in human-robot interactions and discusses a robot's interaction strategies based on gender effects on lexical alignment. We conducted an experiment using a robotic system that interacted with humans in situations where a human referred to an object in an environment and a robot confirmed that indicated object (Fig. 1). Our work is based on past research works about alignment between humans and robots[13][14][16][20].

2. Related Work

In this chapter, we summarize the related works about gender effects on alignment.

2.1 Gender Effects on Alignment between Humans

Some past research has observed differences in the degree of alignment by gender[7]-[9][21]. Namy et al.[8] gave a group of male and female participants a single-word shadowing task to investigate gender differences in vocal alignment. Their experimental results suggest that female shadowers are more likely to align than male shadowers. Levitan et al.[7] measured alignment in three acoustic/prosodic features (intensity, pitch, and jitter) that were extracted from the speech of subjects playing a cooperative computer game and found that alignment is most prevalent for female-male pairs, followed by female-female pairs. The alignment of the male-male pairs was the lowest. Pardo[9] investigated the alignment of pronunciation in task-oriented conversations and found that male talkers overall aligned more than females.

Gender differences in the degree of alignment between humans has also been observed in past work on alignment related to vocal interaction between humans. However, to the best of our knowledge, no research work has investigated lexical alignment. Furthermore, the gender differences in the degree of alignment reported by past research lacks consistency. For example, Namy et al.[8] argued that females are more likely to align than males, in contrast to Pardo's results, where males were more likely to align than females[9].

2.2 Gender Effects on Alignment between Humans and Artificial Media/Robots

Gender differences in the degree of alignment have also been mentioned by past research on human-artificial media/robot interaction[13][15][17]. Thomason et al.[17] investigated the relationships between acoustic/prosodic alignment to a tutoring dialogue system and concluded that males aligned to loudness features more than females. Strupka et al.[15] investigated acoustic/prosodic alignment in human-robot dialogues. Their results showed that the gender of the robot's voice marginally affected the acoustic/prosodic alignment, but they found no human gender effect on it. Iio et al.[13] experimented with a remotely operated robot and investigated whether human pointing was aligned to the robot's gestures. They analyzed their experimental results on participant genders and concluded there is no effect of human gender differences on the alignment of pointing gestures.

In the field of human-artificial media/robot interaction of alignment, gender differences for the degree of alignment have also been discussed. However, to the best of our knowledge, since no past research has treated lexical alignment, we focus on whether gender differences affect the degrees of lexical alignment.

3. Interaction Design

To investigate the gender effects on lexical alignment between humans and robots, we used an interaction called object reference conversations (Fig. 2). Such conversations focus on confirmation behavior, which is often observed in human-human communication. If a person cannot confidently understand which object was being referenced, she is likely to ask for confirmation. Furthermore, people sometimes confirm the referenced object even in case of the referenced object is clear to avoid discrepancies in the interpretation. Such conversations are already being used in human-robot interaction research fields to explore lexical alignment as well as the alignment of pointing gestures in human-robot interaction[13][14][16][20].

Object reference conversations consist of four parts: *Ask*, *Refer*, *Confirm*, and *Answer*. First, a robot asks an interlocutor to refer to an object in an environment (*Ask*). Next the interlocutor refers to an object (*Refer*), and the robot confirms the object to which the interlocutor referred (*Confirm*). Then the interlocutor answers whether the object confirmed by the robot is correct (*Answer*).

In this paper, based on past research works[16][20], we employ two interaction strategies for a robot. The past research works investigated whether lexical alignment occurs in human-robot interaction and its degree. Both strategies are multi-modal interaction considering speech and gestures together. We used the multi-modal strategies rather than strategies that use specific modalities because basically human-human interaction is multi-modal and multi-modal interaction is need for natural interaction between people and robots. A use of only partial modality is also unnatural for interaction with people. Especially, for lexical alignment multi-modal interaction is important because human's gestures are reported to affect lexical alignment. Holler and Wilkin[22] reported that lexical alignment became suppressed when human aligns with their interlocutor's gestures. Iio et al.[14] also suggested that lexical alignment about objects' attributes became suppressed when human align with robot's pointing gesture.

Therefore, we did not investigate the effects of each modality (e.g., gesture only) or different interaction style (e.g., typing) partially; rather we are interested in the gender effects under human-like conversation style, because this style would be common style for social robots which act in real environments. To investigate the gender effects in human-robot interaction research field, we focused on the two major conversation strategies in human-robot interaction under object-reference conversations: explicit and implicit. These conversation strategies are already used to investigate the degree of alignment in human-robot interaction[16][20], therefore using these two strategies would be appropriate for our purpose. The details of each strategy are described as follows.

3.1 Implicit Alignment Strategy One approach is the implicit alignment strategy proposed by Kimoto et al.[16] In this strategy, a robot makes confirmations that contain minimum information for distinguishing objects. Fig. 3 shows an example of an object reference conversation with an implicit alignment strategy.

This strategy exploits alignment in object reference conversations. Based on these three alignment phenomena, lexical alignment, gestural alignment, and alignment inhibition, past work designed robot behavior as follows. A robot should use minimum information for distinguishing among objects in the environment. Alignment inhibition is a formation phenomenon that decreases in some conversations. Through this design, people learn to make references that include sufficient information to identify the objects by reducing the alignment inhibitions. Past work of Kimoto et al.[16] implemented this design in the *Confirm* part in the object reference conversation.

3.2 Explicit Instruction Strategy Another scheme is the explicit instruction strategy. In it, a robot provides instructions about how to refer to objects in a way of instructions that asks the interlocutors to make a reference that includes as much information as possible and requests that they use the information that was missing from the previous references. Fig. 4 shows an example of an object reference conversation with an explicit instruction strategy.

This strategy is based on the following considerations. If an interlocutor refers to an object, as prodded by the robot, it will probably recognize it with high performance. If the interlocutor fails to follow the robot's instructions, the robot should also request that the interlocutor uses all of the instructed information for the following object references. This suggestion encourages the interlocutor to "obey" in subsequent conversations.

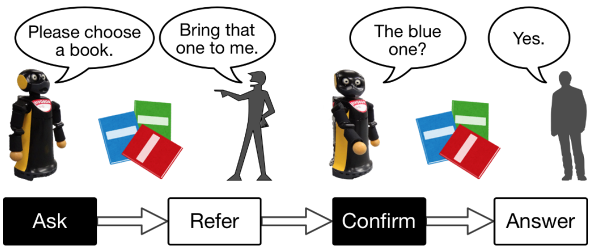

Fig. 2. Object reference conversation: black and white boxes respectively denote robot and human turns

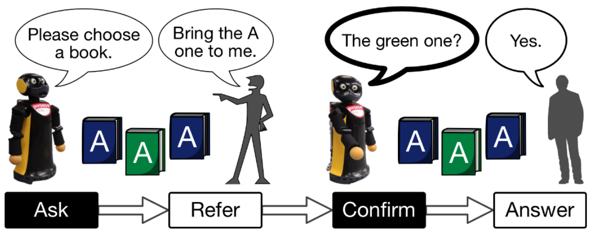

Fig. 3. Example of object reference conversation with implicit alignment approach: robot confirmed indicated object using minimum information for distinguishing objects (thick line denotes the part implicit alignment strategy is implemented in)

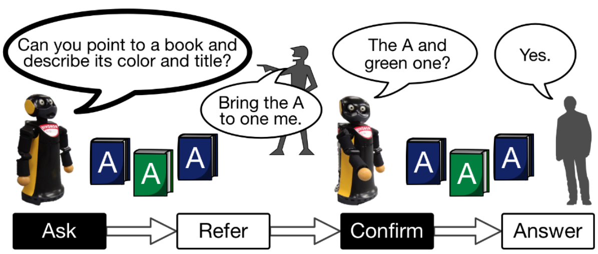

Fig. 4. Example of explicit robot instructions: robot explicitly instructs how to refer to objects (thick line denotes the part explicit instruction strategy is implemented in)

4. System

We developed a system based on past works that implemented implicit alignment and/or explicit instructions for object reference conversations[16][20]. The system consists of four parts: sensors, an indicated object recognition function, an object information database, and a robot behavior control function. When a user refers to an object, the indicated object recognition function identifies the user's reference behavior and estimates the indicated object. The robot behavior control function chooses a robot behavior that corresponds to the implemented strategy and sends a behavior command to the robot, which confirms the indicated object and asks a user to refer to it in the next conversation in a manner decided by the robot behavior control function. Fig. 5 shows the architecture of our developed system.

The system can also have object reference conversations as a basic function. In its *Ask* and *Confirm* parts, the robot performs a behavior that corresponds to whichever approach was used by the robot.

4.1 Robot In this study, we used Robovie-R ver.2, a humanoid robot developed by the Intelligent Robotics and Communication Labs, ATR, which has a human-like upper body designed for communication with humans. It has three DOFs for its neck and four for each arm. Its body has sufficient expressive ability for object reference conversations. We used XIMERA for

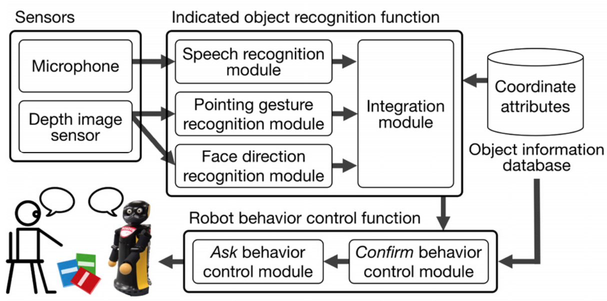

Fig. 5. System architecture to recognize objects indicated by interlocutor

speech synthesis[23]. It is 1100-mm tall, 560-mm wide, 500-mm deep, and weighs about 57 kg.

4.2 Indicated Object Recognition Function To develop this function, we implemented an algorithm[16][20] that combines the speech recognition, pointing gesture recognition, and face direction recognition results.

4.2.1 Speech Recognition Module The speech recognition module receives human speech that refers to an object and outputs the normalized reference likelihood of each object based on speech recognition. To calculate the likelihood, we used the number of attributes in human speech[16], which were captured by a microphone attached to a human's collar. In this system, we used a speech recognition engine called Julius, which gives good performance for Japanese[24].

4.2.2 Pointing Gesture Recognition Module The pointing gesture recognition module obtains the body frame data from a depth image sensor called Kinect for Windows v2 and outputs the normalized reference likelihood of each object based on pointing gesture recognition. We modeled the likelihood as the difference from the pointing vector (between the human head and the tip of the human hand) to a vector between the human head and an object with a normal distribution function $N(0, 1)$.

4.2.3 Face Direction Recognition Module The face direction recognition module obtains the face direction vector from the depth image sensor and outputs the reference likelihood based on the face direction recognition. We modeled the likelihood based on an angle parallel to the plane of the floor between the face direction vector and a vector between a human head and an object. If the vector is less than 110º, the person is considered to be viewing the object; its likelihood is 1, and otherwise 0. This is because a human's field of view is 110º at most[25]. The likelihoods are finally normalized from 0 to 1.

4.2.4 Integration Module The integration module merges the reference likelihoods of the speech and both the pointing gesture and face direction recognitions. These three likelihoods are summed and normalized[16]. The object with the highest likelihood is estimated to be the one indicated by the interlocutor.

4.3 Robot Behavior Control Function The robot behavior control function determines how the robot confirms the indicated object (*Confirm* behavior) and how it asks an interlocutor to refer to it (*Ask* behavior) in subsequent conversations. The conversation contents of the *Confirm* and *Ask* behaviors reflect whether the implicit alignment strategy or the explicit instruction strategy is used.

When using the implicit alignment strategy, this function chooses the *Confirm* behavior and adopts the implicit alignment strategy, and the robot confirms the indicated object with minimum information for distinguishing among objects. The *Ask* behavior does not adopt a particular strategy, and the robot does not explicitly instruct the participants how to make references.

When using the explicit instruction strategy, this function chooses the *Ask* behavior and adopts the explicit instruction approach, and the robot explicitly provides instructions about how to refer to objects. The *Confirm* behavior does not adopt a particular strategy, and the robot confirms the indicated object by pointing and verifying all of the information about it.

5. Experiment

5.1 Hypotheses and Predictions Some past research works reported a gender effect on alignment. However, such gender differences reported by past research are inconsistent. One reported tendency is that females align more than males[7][8], but another argues the opposite[9][17]. Since predicting the tendency related to gender effects on alignment is difficult, we made two contradictory hypotheses about gender effects on lexical alignment in human-robot interaction.

Hypothesis about female-dominant effects on lexical alignment in human-robot interaction

Past research identified female-dominant effects on alignment. Namy et al.[8] reported that females are more likely to align than males even when social interaction is severely limited. Their participants did a shadowing task in which they sat alone in a room and repeated single words uttered by various speakers over headphones. Levitan et al.[7] measured alignment on acoustic/prosodic by analyzing the speech of participants who were playing cooperative computer games for female-female, female-male, and male-male dyads and found that alignment is most prevalent for female-male pairs, followed by female-female pairs. Male-male pairs aligned the least. Although their results didn't show that female-female pairs aligned the most, the least aligned pairs were the male-male pairs. These results suggest that females align more than males. Therefore, we believe that females will align more with interlocutors than males. Based on these considerations, we made the following hypothesis:

Prediction 1-a: Females will lexically align more with a robot interlocutor than males.

Hypothesis about male-dominant effects on lexical alignment in human-robot interaction

Past research identified male-dominant effects on alignment. Pardo[9] concluded that the speech of talkers became more similar to the pronunciation of their partner's speech during conversational interactions. She reported that overall, male talkers were more aligned than female talkers. Thomason et al.[17] investigated whether students acoustically/prosodically aligned with a tutoring dialogue system. Each student verbally responded to either pre-recorded or synthesized tutor questions. Their results suggested that males were significantly more aligned than females to minimum and maximum features of loudness. Therefore, we believe that males will align more with interlocutors than females. Based on these considerations, we made the following hypothesis:

Prediction 1-b: Males will lexically align more with a robot interlocutor than females.

5.2 Environment Our participants sat in front of the robot. We arranged books as objects in the environment by following past works that used object reference conversations[13][14][16][20]. Five books were placed in a 1.5 m by 3.3 m rectangular area between the robot and the participant and grouped close together without overlapping approximately 0.6 – 2.6 m from the participants. Fig. 6 shows the experimental environment.

We controlled the attributes of the books based on past research work that focused on object reference conversations[16][20]. All the books were 21 cm by 27.5 cm, and their attributes were color, symbol, and a letter on the cover. There were three colors: red, blue, or yellow. Three symbols were placed on the book covers: a circle, a triangle, or a square. There were two letters: Q and B. We prepared 18 books to satisfy all combinations of attributes.

5.3 Conditions We controlled the strategy that was applied to our developed system (applied strategy factor). The applied strategy factor had two levels: implicit alignment and explicit instruction. Both were applied to the *Confirm* and *Ask*

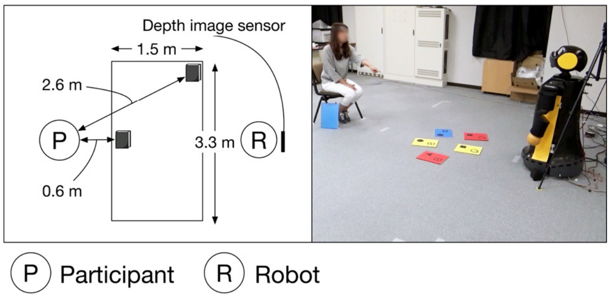

Fig. 6. Experimental environment

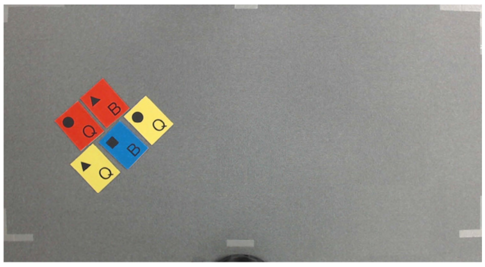

Fig. 7. Example of book arrangement

parts of the object reference conversations. The applied strategy factor had a within-participant condition. There was no difference in the manner of recognizing the interlocutor's reference behavior or estimating the indicated object.

5.3.1 Implicit Alignment Condition In the implicit alignment condition, unlike the explicit instruction condition, the robot did not explicitly provide instructions about the reference style; it just said, "Please choose a book" in the *Ask* part.

On the other hand, in the *Confirm* part the robot said a different sentence. For this purpose, we implemented an implicit alignment design for the reference behavior. In this condition, the robot confirmed the object with minimum information for distinguishing among objects; confirmations were based on the implicit alignment strategy of references. This approach determined the robot's object reference behaviors, i.e., with or without a pointing behavior and speech contents, by considering the objects' position relationships and characteristics. The robot pointed to reduce the number of candidates of the referenced objects. The speech format of the confirmations is the sequence of object attributes. For example, the robot asks, "That yellow book with a triangle on its cover?" or "That blue book?"

5.3.2 Explicit Instruction Condition In the explicit instruction condition, the robot gives instructions about how to refer to objects in a way that asks interlocutors to make a reference that includes as much information as possible in the *Ask* part.

The speech format of the explicit instructions includes two sentences. The first is used every time; the second is only used when a participant failed to use all of the information requested in the first sentence in the previous reference.

For example, the robot asks, "Can you refer to the book by color, the symbol and the letter on its cover as well as by pointing and looking at it? Please refer to a letter and point."

In the *Confirm* part of this condition, since the robot verified the objects with all of the information, it gave every attribute of an object and pointed during the confirmations.

5.4 Participants Twenty people (ten females and ten males who averaged 35.5 years of age, $SD=9.9$) participated in our experiment. We decided the number of subjects based on past research works about alignment. Five seventh of past research works we cited in chapter "2. Related Work" investigated gender effects on alignment by less than 20 subjects except the number of subjects who assesses alignment: Levitan et al.(7), Namy et al.(8), Pardo(9), Iio et al.(13) and Strupka et al.(15) For example, in the Namy et al.'s shadowing task 8 female and 8 male shadowers repeated words sounded from headphones(7). In the Iio et al.'s task 10 female and eight male participated the conversation with the robot(13). Although experimental procedures of the five past research works are all different and comparing the number of subjects simply is difficult, both number of female and male subjects is less than or equal to 10(7)-(9)(13)(15).

5.5 Procedure We conducted our experiment as follows. First, we explained it to the participants who signed consent forms. Next, we orally gave them the following instructions: "The robot can recognize human speech, pointing gestures, and face directions. It will ask you to indicate a book. Do so as if you were dealing with a person".

After the instructions, the participants selected five books from among the 18 and arranged them based on the experimenter's instruction: "Please arrange the books in one place". An example of the arrangement is shown in Fig. 7. After that the participant repeated the object reference conversations ten times. We call this set of ten object references the conversation sessions, which were conducted in both the applied approach conditions: explicit instruction and implicit alignment. The participants answered questionnaires about their impressions of the conversations after each conversation session. We counterbalanced the order of the interactive strategy conditions.

5.6 Measurement

5.6.1 Information Amount of Reference To investigate the change of reference styles, we measured the mean number of object attributes (color, symbol, and letter) in the participant references per sessions.

According to lexical alignment findings between humans and robots, humans tend to use the same word as the robot in conversations(14). This finding suggests, for example, if a robot uses the word "blue" as a color attribute, humans will avoid the word "cyan" and use blue instead. Therefore, we measured the mean number of object attributes contained in the robot's object information database and used in the *Confirm* part.

5.6.2 Reference Redundancy of Utterances The reference redundancy of an utterance is defined as the difference between the numbers of object attributes in the participant's references and the minimum number of attributes for uniquely identifying the referenced objects in the environment per sessions. Our objects have three attributes (color, symbol, and letter), and the number of object attributes in the participant's references was defined as 0 to 3. For example, if a participant's reference has no attributes, the numbers of object attributes is 0. If a participant's reference has all three attributes (color, symbol, and letter), the numbers of object attributes is 3. The minimum number of attributes to uniquely identify the indicated object in the environment ranges from 1 to 3. For these reasons, reference redundancy ranges from −3 to 2. For example, if a participant refers to a book with no attributes (i.e., "that book" or "this book") in the environment where all the attributes are needed to uniquely identify objects (minimum number of attributes is 3), the reference redundancy of utterances is −3. We measured the reference redundancy of the utterances for the following two reasons.

First, lexical alignment does not just increase the use of the word contained in the robot's attributes; it also leads to alignment of word selection/combination. For example, if a robot uses "blue and B" when it refers to objects, humans tend to use the same selection/combination of words (color and symbol). Such alignment, called word selection/combination, is also observed in human-robot interaction[(14)(26)]. For our experiment, the robot selects words based on two strategies: implicit alignment and explicit instruction. As mentioned in Section 5.3, in both strategies, the robot uses a word combination that can uniquely identify the objects referenced in the environment. With such lexical alignment, humans tend to use words that can identify objects. For example, in an environment that only has red books, humans rarely use "That red book" as the reference, but instead they say "That red book with a circle and a B on its cover" because the second reference way clearly identifies the object in the environment.

Second, objects in environments differ with respect to conversation sessions and participants, and value of the number of object attributes in a participant's references depends on the environment. For example, the value of "red" as a color attribute in an environment that only has red books is much less than its value in an environment that has red, blue, and yellow books.

6. Results: Verification of Prediction 1

Fig. 8 shows the results of the information amount of references. We conducted a two-factor mixed ANOVA for both applied strategy and gender factors and identified significant main effects in the applied strategy factor ($F(1,18) = 6.616$, $p = .019$, partial $\eta^2 = .269$). We found no significance in the gender factor ($F(1,18) = 2.646$, $p = .121$, partial $\eta^2 = .128$) and no significant interaction ($F(1,18) = .952$, $p = .342$, partial $\eta^2 = .050$). These results showed that the number of object attributes in the references with explicit instruction was significantly larger than the number of object attributes in references with implicit alignment. On the other hand, these results showed no gender-based differences in the information amount of the references.

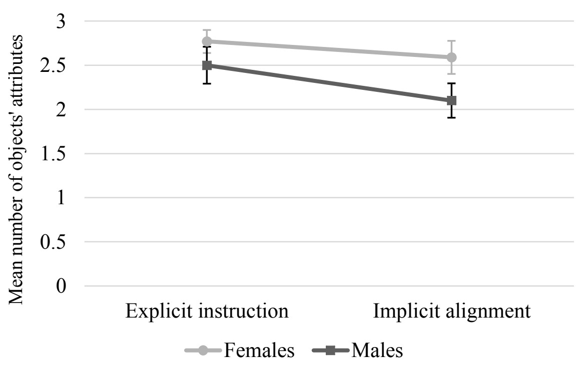

Fig. 8. Information amount of reference

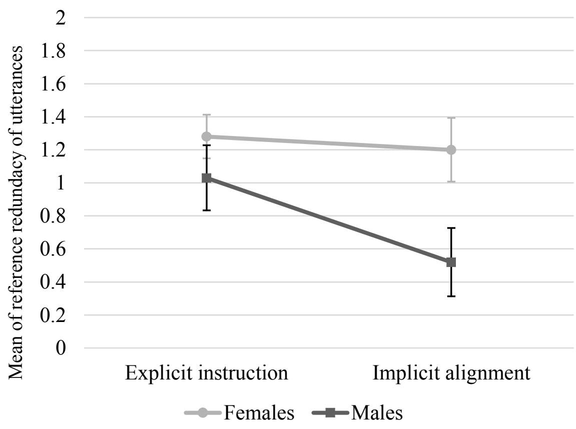

Fig. 9. Reference redundancy of utterance

Fig. 9 shows the results of the reference redundancy of utterances. We conducted a two-factor mixed ANOVA for both applied strategy and gender factors and found significant main effects in the applied strategy factor ($F(1,18) = 4.485$, $p = .048$, partial $\eta^2 = .199$) and gender factor ($F(1,18) = 4.423$, $p = .050$, partial $\eta^2 = .197$); we found no significant interaction ($F(1,18) = 2.382$, $p = .140$, partial $\eta^2 = .117$). These results showed that reference redundancy with explicit instruction was significantly larger than reference redundancy with implicit alignment. They also showed that the reference redundancy of females exceeds that of males.

From these results on the amount of information about references and the reference redundancy of utterances, prediction 1-a is partially supported, but prediction 1-b is not supported.

7. Discussion

7.1 Implication Our experimental results showed that females refer to objects with references that have higher reference redundancy of utterances than males. The reference redundancy of utterances reflects how useful references are for identifying objects, and our experimental results suggest that robots need to change their interaction strategies for effective alignment with human references to useful references to identify objects in object reference conversations. For example, when a robot uses the implicit alignment strategy, the reference redundancy of male utterances is relatively lower than that of females, and therefore robots should choose the explicit instruction strategy to obtain useful information to identify objects from males. Since the overall conversation impressions did not differ by gender or applied strategy factors, explicit instruction to males by robots would have fewer disadvantages. For these reasons, considering gender effects on lexical alignment is important for designing conversation strategies for robots. Our findings might be integrated not only for object reference conversation contexts but also for other conversation contexts, since lexical alignment is not a phenomenon that is only observed in object reference conversations.

For the information amount, there were no gender effects on lexical alignment, although there were gender differences for the reference redundancy of utterances. This result suggests that even though males aligned as many words as females, they aligned with fewer word combinations that can uniquely identify the referenced object in the environment than females. Namy et al.[(8)] found that in shadowing tasks, females were vocally more likely to align than males. They discussed why females aligned more than males, and posited females might be more sensitive to the indexical features of interlocutors. If sensitivity to conversational features differs by gender based on their discussions, the difference of sensitivity might explain the discrepancy of our results between information amount and reference redundancy.

We found significant main effects in the gender factor about the reference redundancy of utterances ($F(1,18) = 4.423$, $p = .050$, partial $\eta^2 = .197$). Although the interpretation of effect size varies

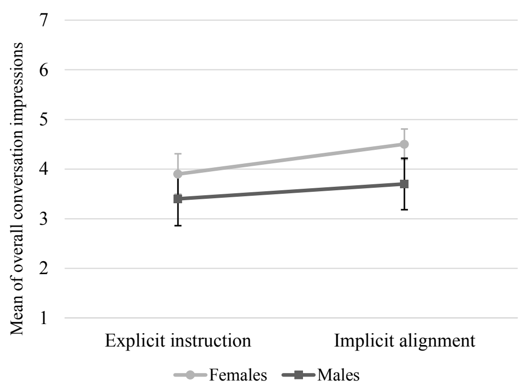

Fig. 10. Overall conversation impression

by experiments, Cohen[27] offers standard interpretation of partial η^2 as benchmarks: small, medium and large effects would be reflected in partial $\eta^2 = .0099$, .0588 and .1379 respectively. Compared to the Cohen's benchmark, the gender factor has large effects on the reference redundancy of utterances.

7.2 Conversation Impressions To investigate the participant's impressions of the conversations, we measured a questionnaire item, overall conversation impression of the robot, and evaluated it on a 1-to-7 point scale. Fig. 10 shows the questionnaire results about the overall conversation impressions. We conducted a two-factor mixed ANOVA for both factors, applied strategy and gender, and found no significance in the applied strategy factor ($F(1,18) = 2.751$, $p = .115$, partial $\eta^2 = .133$), no significance in the gender factor ($F(1,18) = 1.254$, $p = .278$, partial $\eta^2 = .065$) and no significant interaction ($F(1,18) = .306$, $p = .587$ partial $\eta^2 = .017$).

These results show that overall conversation impressions did not differ based on the applied strategies and genders.

7.3 Limitations We conducted our experiment's study in a limited situation. The participants referred to objects with only three features: color, a symbol, and a letter. In real environments, the features of objects are not limited and obviously influence the reference ways. But since the interaction manner between a robot and an interlocutor does not depend on features, our findings are general for other objects.

Since our experiment was conducted with an existing robot named Robovie-R ver.2, robot generality is also limited. Some past research works on gender effects on lexical alignment investigated from the viewpoint of gender pairs[7][9][15]. Robovie-R ver.2 and its synthesized speech have no intended gender. If a robot and/or its speech are designed to represent a specific gender, the robot's gender will undoubtedly also influence lexical alignment in conversations.

8. Conclusion

We investigated gender effects on lexical alignment in object reference conversation contexts between humans and robots by employing two interaction strategies based on related works: implicit alignment and explicit instruction. We developed a system that recognized the indicated objects and did object reference conversations with humans. Experimental results indicated that females lexically align more with a robot interlocutor than males in terms of reference redundancy of utterances. Our female participants aligned more than males and used more references that are useful to uniquely identify referenced objects in the environment. We believe that our findings of the female-dominant effects on lexical alignment in human-robot interaction will help robotics researchers design conversation strategies between humans and robots.

Acknowledgements

This research work was supported by JSPS KAKENHI Grant Numbers JP15H05322, JP16K12505, and JP15K16075.

References

(1) M. Kimoto, T. Iio, M. Shiomi, I. Tanev, K. Shimohara, and N. Hagita : "Relationship between Personality and Robots' Interaction Strategies in Object Reference Conversations", The Second International Conference on Electronics and Software Science, pp.128-136, Takamatsu, Japan (2016)

(2) S. E. Brennan and H. H. Clark : "Conceptual pacts and lexical choice in conversation", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol.22, No.6, pp.1482-1493 (1996)

(3) S. E. Brennan : "Lexical entrainment in spontaneous dialog", 1996 International Symposium on Spoken Dialogue, pp.41-44 (1996)

(4) A. Nenkova, A. Gravano, and J. Hirschberg : "High frequency word entrainment in spoken dialogue", The 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers, pp.169-172, Columbus, Ohio, USA (2008)

(5) C. C. Lee, M. Black, A. Katsamanis, A. C. Lammert, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan : "Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples", INTERSPEECH 2010, pp.793-796 (2010)

(6) M. J. Pickering and S. Garrod : "Toward a mechanistic psychology of dialogue", *Behavioral and Brain Sciences*, Vol.27, No.2, pp.169-190 (2004)

(7) R. Levitan, A. Gravano, L. Willson, S. Benus, J. Hirschberg, and A. Nenkova : "Acoustic-prosodic entrainment and social behavior", The 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp.11-19, Montréal, Canada (2012)

(8) L. L. Namy, L. C. Nygaard, and D. Sauerteig : "Gender Differences in Vocal Accommodation", *Journal of Language and Social Psychology*, Vol.21, No.4, pp.422-432 (2002)

(9) J. S. Pardo : "On phonetic convergence during conversational interaction", *The Journal of the Acoustical Society of America*, Vol.119, No.4, pp.2382-2393 (2006)

(10) J. Pearson, J. Hu, H. P. Branigan, M. J. Pickering, and C. I. Nass : "Adaptive language behavior in HCI: how expectations and beliefs about a system affect users' word choice", SIGCHI Conference on Human Factors in Computing Systems 2006, pp.1177-1180, Montréal, Canada (2006)

(11) H. P. Branigan, M. J. Pickering, J. Pearson, and J. F. McLean : "Linguistic alignment between people and computers", *Journal of Pragmatics*, Vol.42, No.9, pp.2355-2368 (2010)

(12) H. P. Branigan, M. J. Pickering, J. Pearson, J. F. McLean, and C. Nass : "Syntactic alignment between computers and people: The role of belief about mental states", 25th Annual Conference of the Cognitive Science Society, pp.186-191 (2003)

(13) T. Iio, M. Shiomi, K. Shinozawa, T. Akimoto, K. Shimohara, and N. Hagita : "Investigating Entrainment of People's Pointing Gestures by Robot's Gestures Using a WOZ Method", *International Journal of Social Robotics*, Vol.3, No.4, pp.405-414 (2011)

(14) T. Iio, M. Shiomi, K. Shinozawa, K. Shimohara, M. Miki, and N. Hagita : "Lexical Entrainment in Human Robot Interaction", *International Journal of Social Robotics*, Vol.7, No.2, pp.253-263 (2015)

(15) E. Strupka, O. Niebuhr, and K. Fischer : "Influence of Robot Gender and Speaker Gender on Prosodic Entrainment in HRI", Interactive Session at 2016 25th IEEE International Symposium on Robot and Human Interactive Communication, New York City, USA (2016)

(16) M. Kimoto, T. Iio, M. Shiomi, I. Tanev, K. Shimohara, and N. Hagita : "Improvement of object reference recognition through human robot alignment", 2015 24th IEEE International Symposium on Robot and Human Interactive Communication, pp.337-342 (2015)

(17) J. Thomason, H. V. Nguyen, and D. Litman : Prosodic Entrainment and Tutoring Dialogue Success, In: H. C. Lane, K. Yacef, J. Mostow, and P. Pavlik, editors, Artificial Intelligence in Education, pp.750-753, Springer Berlin Heidelberg, Berlin, Heidelberg (2013)

(18) J. A. Hall : "Gender effects in decoding nonverbal cues", *Psychological bulletin*, Vol.85, No.4, pp.845-857 (1978)

(19) J. S. Hyde : "Gender Similarities and Differences", *Annual Review of Psychology*, Vol.65, pp.373-398 (2014)

(20) M. Kimoto, T. Iio, M. Shiomi, I. Tanev, K. Shimohara, and N. Hagita : "Alignment Approach Comparison between Implicit and Explicit Suggestions in Object Reference Conversations", The Fourth International Conference on Human Agent Interaction, pp.193-200, Biopolis, Singapore (2016)
(21) Z. Xia, R. Levitan, and J. Hirschberg : "Prosodic entrainment in Mandarin and English: A cross-linguistic comparison", 7th Speech Prosody (2014)
(22) J. Holler and K. Wilkin : "Co-Speech Gesture Mimicry in the Process of Collaborative Referring During Face-to-Face Dialogue", *Journal of Nonverbal Behavior*, Vol.35, No.2, pp.133-153 (2011)
(23) H. Kawai, T. Toda, J. Ni, M. Tsuzaki, and K. Tokuda : "XIMERA: A new TTS from ATR based on corpus-based technologies", Fifth ISCA Workshop on Speech Synthesis (2004)
(24) A. Lee, T. Kawahara, and K. Shikano : "Julius — an open source real-time large vocabulary recognition engine", 7th European Conference on Speech Communication and Technology, pp.1691-1694, Aalborg, Denmark (2001)
(25) S. J. Ryan, A. P. Schachat, C. P. Wilkinson, D. R. Hinton, S. V. R. Sadda, and P. Wiedemann : Retina, Elsevier Health Sciences (2012)
(26) S. Kazuhiko, M. Takahiro, K. Masayuki, and N. Hagita : "User specification method and humanoid confirmation behavior", 2007 7th IEEE-RAS International Conference on Humanoid Robots, pp.366-370 (2007)
(27) J. Cohen : Statistical power analysis for the behavioral sciences, Academic Press, New York (1969)

Mitsuhiko Kimoto

(Non-member) received his M. Eng. degree from Doshisha University, Kyoto, Japan, in 2016. He is currently a Ph.D. student at Doshisha University and a Student Intern at the Intelligent Robotics and Communication Laboratories (IRC), Advanced Telecommunications Research Institute International (ATR), Kyoto, Japan. His research interests include human-robot interaction.

Takamasa Iio

(Non-member) received his M. Eng. and Ph.D. degree from Doshisha University, Kyoto, Japan, in 2009 and 2012. He is currently an assistant professor at Osaka University, Osaka, Japan. His research interests include human-robot interaction, group conversation between humans and multiple robots and social behaviors of robots.

Masahiro Shiomi

(Non-member) received M. Eng. and Ph.D. degrees in engineering from Osaka University in 2004 and 2007. From 2004 to 2007, he was an intern researcher at the Intelligent Robotics and Communication Laboratories (IRC). He is currently a group leader in the Agent Interaction Design department at IRC, Advanced Telecommunications Research Institute International (ATR). His research interests include human-robot interaction, robotics for child-care, networked robots, and field trials.

Ivan Tanev

(Non-member) was born in 1964 in Simeonovgrad, Bulgaria. He earned M.S. (with honors) and Ph.D. degrees from Saint-Petersburg State Electrotechnical University, Russia in 1987 and 1993 respectively, and Dr.Eng. degree from Muroran Institute of Technology, Japan in 2001. He has been with the Bulgarian Space Research Institute (1987), Bulgarian Central Institute of Computer Engineering and Computer Technologies (1988-1989), Bulgarian National Electricity Company (1994-1997), Synthetic Planning Industry Co. Ltd., Japan (2001-2002), and ATR Human Information Science Laboratories (2002-2004), Japan. Since April 2013 he has been a professor at Doshisha University, Japan. Dr. Tanev's research interests include evolutionary computations, evolutionary robotics and multi-agent systems.

Katsunori Shimohara

(Member) received the B.E. and M.E. degrees in Computer Science and Communication Engineering and the Doctor of Engineering degree from Kyushu University, Fukuoka, Japan, in 1976, 1978 and 2000, respectively. He was Director of the Network Informatics Laboratories and the Human Information Science Laboratories, Advanced Telecommunications Research Institute (ATR) International, Kyoto, Japan. He is currently a Professor at the Department of Information Systems Design, Faculty of Science and Engineering, ant the Graduate School of Science and Engineering, Doshisha University, Kyoto, Japan. His research interests include human communication mechanisms, evolutionary systems, human-system interactions, and socio-informatics.

Norihiro Hagita

(Non-member) received B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University in 1976, 1978, and 1986. In 1978, he joined Nippon Telegraph and Telephone Public Corporation (now NTT), where he developed handwritten character recognition. He was a visiting researcher in the Department of Psychology, University of California, Berkeley from 1989-90. He is currently a Board Director of ATR and an ATR Fellow, a director of the Social Media Research Laboratory Group and Intelligent Robotics and Communication laboratories. His research interests include networked robotics, human-robot interaction, pattern recognition, and data-mining technology. He is a member of IEEE, the Information Processing Society of Japan, and the Japanese Society for Artificial Intelligence, The Institute of Electronics, Information and Communication Engineers (IEICE) and The Robotics Society of Japan.