International Journal of Social Robotics (2020) 12:1021–1029
https://doi.org/10.1007/s12369-017-0462-3

Language Use in Joint Action: The Means of Referring Expressions

Harumi Kobayashi[1] · Tetsuya Yasuda[2] · Hiroshi Igarashi[3] · Satoshi Suzuki[4]

Accepted: 26 December 2017 / Published online: 13 January 2018

Abstract
This study examined how human–human collaboration can be achieved through an exchange of verbal information in exchanging information about the referents in a joint action. Knowing other people's referential intention is fundamental for joint action. Joint action can be achieved verbally by two types of referring expressions, namely, symbolic and deictic referring expressions. Using corpus data, we extracted nouns as typical symbolic references and demonstratives as typical deictic references. We examined whether the word usage of these terms changed when the robot vehicles controlled by the participants repeatedly performed the same collaborative task. We used a novel virtual space for the task because we wanted to control the common ground shared by the participants. The results of the performance indicate that the task completion became more efficient as the participants repeated the task. The referential word use was reduced in both symbolic and deictic references, and this reduction occurred with a grounding process among the collaborators. The study showed that reduction of referential expressions occurs with the grounding process in human–human collaboration and suggests that appropriate collaborative robot systems must deal with the reduction process of referencing in humans.

Keywords Collaborative work · Joint action · Grounding · Common ground · Demonstratives · Referring expressions

1 Introduction

In the present highly industrialized societies, robots have been operational in factory and manufacturing settings as well as in natural human environments, such as homes, stores, hospitals, and museums. However, a majority of these robots can only function with predetermined programs or through remote control by humans. Autonomous robots that can work with humans may be ideal in conditions that often involve unpredictable situations. However, perfect autonomous robots are yet to be realized because the possible mechanisms of human–robot collaboration are not well known. This study explores the information exchange on the referents in human–human collaboration, hoping to provide some insight into the design of human–robot information exchange in joint action.

Cooperation is regarded as natural human behavior [1]. When there is a concerted effort in collaborative situations, such as in moving objects, people act jointly. Sebanz et al. [2] defined joint action as "any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment" (p. 70). Furthermore, Sebanz et al. [2] discussed three important components in joint action: joint attention, task sharing, and action coordination.

Language can be a powerful tool to solve problems that arise in joint action. For example, in joint attention, the other person's referential intention is usually described with referential words. Distinctive referential expressions used in verbal information include the following: (a) nouns, such as common nouns (e.g., box, car) and proper nouns (e.g., David, Kobayashi-san), (b) deictic words such as demonstratives (e.g., this, that, here, there), and (c) pronouns (e.g., he, she, it).

A growing body of research has focused on the use of language in collaboration, mostly on speaker–addressee pairs, using referential communication game tasks. The findings showed several important facts on collaborative language use

✉ Harumi Kobayashi
h-koba@mail.dendai.ac.jp

[1] Division of Information System Design, Tokyo Denki University, Saitama, Japan

[2] Department of Human Developmental Psychology, Jumonji University, Saitama, Japan

[3] Department of Electrical and Electronic Engineering, Tokyo Denki University, Tokyo, Japan

[4] Department of Robotics and Mechatronics, Tokyo Denki University, Tokyo, Japan

[3–5]. First, the speaker adjusts the language use by considering the addressee's knowledge of the task. One clear example is that the speaker provides newly given information to the addressee using the indefinite article "a," such as "a diamond [6]," for a new referent in the discourse ("there [is] a diamond"). But in the repeated references to the same object, the speaker uses the definite article "the" to indicate that the referent is a shared known object ("right under the diamond…"). Second, the speaker shows a tendency to use more words to describe the referent for the first time. For example, the speaker first mentions "a figure…something like a monk praying." However, in the repeated referencing, the speaker uses reduced forms such as "the monk praying," or "the monk" [3]. This necessarily reduces the number of used words. Third, these language changes seem to occur through a "grounding" process. Grounding means the interpretation of any linguistic expression by considering a previously shared common ground [7–9]. This process includes (a) the discourse history, (b) the ongoing discourse between the interlocutors, and (c) the shared knowledge between the interlocutors as members of certain social groups.

Based on the literature review of language use in joint action, we address three issues: the grounding process, reference words, and joint action by more than two people. First, the actual grounding process is yet to be clarified because the types and amount of common ground shared by interlocutors are difficult to assess. Kobayashi et al. [10] examined the conversation of three people in joint action in a virtual space. They focused on verb use because the task required the repositioning of boxes. Additionally, they had anticipated that verb use would change when the task was repeated. The result confirmed that the number of verb types decreased and reduced to a few verbs, such as "push" and "stop." Based on the results, they inferred that the establishment of common ground among the participants might have contributed to this reduction of verb types. However, they did not generate relevant data on the construction and sharing of a common ground; therefore, the relationship between common ground and verb use could not be analyzed.

Second, how a target of joint action is referred to has not been thoroughly explored yet. This is unfortunate, because specifying and sharing targets is fundamental in joint action. It has been suggested that deictic referring such as eye fixation and pointing are a very effective and efficient means for human–robot interaction. Ballard et al. [11] and Kooijmans et al. [12] illustrated that the use of deictic referring can be more effective than the specification of time and space. Sato et al. [13] also discussed the usefulness of deictic gestures in a human–robot interaction. Using a highly specified situation, Sugiyama et al. [14] showed that the deictic referring (i.e., the use of eye fixation and demonstratives) could have been more effective than the symbolic referring (i.e., the use of numbers) in a human–robot interaction.

These studies, however, did not examine the different natures of deictic reference and symbolic reference and their relation to the grounding process. In addition, we can state that nouns can be replaced with demonstratives, so reduction of these two types may occur differently. For example, "the box" may be replaced with "this" in the next reference when the referent is already shared. Thus, the use of demonstratives may be more frequent when a task is repeated. If the use of deictic reference is more efficient than that of symbolic reference in human–robot interaction, the reduction of deictic reference may not be evident when compared with the reduction of symbolic reference.

Third, as most of the previous research studies on joint action examined speaker–addressee pairs, language changes that may occur when more than two people are involved are still unexplored. The involvement of a third person or more participants complicates the task situation further. However, group work by more than two people is common in human–human collaboration.

In this study, we examine how three people exchange information about referents in a discourse when they repeat the same collaborative task of moving objects in a novel environment. Among the referential expressions, we focus on the use of common nouns and demonstratives as typical reference words. These two category terms have different features. Common nouns are generic words with meanings that can be naturally understood without context (i.e., the meaning of the word "box" is not ambiguous) [15]. However, demonstratives are context-bound, and the actual referent in a given situation may change (e.g., the meaning of the word "this" depends on the situation) [16]. In addition, because demonstratives have fewer syllables, they can be quickly pronounced in all human languages [16].

We used a novel task and environment by constructing a virtual task field on the computer monitor. The reason is that we intended to control the participants' initial common ground in terms of the knowledge about the task. We used two objective measures, namely, task completion time, and ratio of robot movement and object movement. By doing so, we intended to estimate the level of grounding. We asked the participants to complete the same task ten times to observe whether the use of referential expressions may change in repeated collaborative experience. We extracted nouns and demonstratives and examined whether the use of these words was reduced in repeated referencing.

As for nouns, we further categorized the extracted nouns into groups of (a) common nouns (e.g., box, car, wall), (b) direction nouns (e.g. right, left), (c) space nouns (e.g., corner, "sukima" [narrow space in Japanese]), and (d) time nouns (e.g., "ima" [now], "ato" [later]). We intended to observe whether the distribution of these nouns changes over repeated task trials. In Japanese, time nouns convey the same meaning as English adverbs of time. Japanese time nouns become

adverbs denoting time if particles are added, such as "ato-ni" (later). We extracted the nouns of direction, space, and time. The reason is that these "indices" must satisfy the location of the objects in space and in time. Furthermore, these nouns may show other requirements, such as the object description and the speaker's presuppositions of the addressee's knowledge [7].

Demonstratives such as "this" and "that" are a unique class of deictic words that play an important role in joint action [14]. When a person says, "I want to buy this clock," he/she obviously intends to buy a specific clock and believes that the addressee would know the correct referent compared to her saying "I want to buy a clock." Based on Clark's [7] description of the importance of demonstrative referring, Diessel [17] suggested that, "while there are many linguistic means that speakers can use to coordinate a joint attentional focus, there is no other linguistic device that is so closely tied to this function than demonstratives" (p. 469). Demonstratives are primitive and short in linguistic forms, and observed in all human languages [9,16]. The use of demonstratives is usually based on the distance to the target object [16–19]. On Japanese demonstratives, Takahashi and Suzuki [20] and Endo [21] examined the effects of distance to the referent in the use of the demonstrative pronoun "*Kore*" (This) and the demonstrative adjective "*Kono*" (This), "*Sore*" (demonstrative pronoun), and "*Sono*" (demonstrative adjective) (hereafter named as the *That-proximal*), and "*Are*" (demonstrative pronoun) and "*Ano*" (demonstrative adjective) (hereafter named as the *That-distal*) [16,17,20,21]. They revealed the following characteristics:

1. The demonstrative "This" was used when the speaker's position was proximate to the target.
2. The demonstrative "That-proximal" was used when the speaker's position was intermediate to the target or when the hearer's position was proximate to the target.
3. The demonstrative "That-distal" was used when the speaker's position and the hearer's position were both far from the target.

We examined all the demonstratives and nouns used in the four groups to analyze for any change in proficiency of these words during the task. First, we anticipated that the reduction of words would occur as grounding proceeds. Second, the reduction would be evident in common nouns but not evident in task-specific nouns (i.e., direction, space, and time nouns). While some of the common nouns may be replaced by the demonstrative pronouns, namely, "this" or "that," the task execution requires the use of a certain number of task-specific nouns. Third, the reduction of demonstratives would also occur concurrently with the reduction of common nouns, but the frequency in certain demonstratives would remain unchanged. The reason is that demonstratives are generally used to control joint action. Fourth, these changes of language use would correlate with grounding. Here, we computed the estimated level of grounding with two measures: the task completion time and the efficiency of the robot travelling distance.

The efficiency of the robot's travelling was calculated by the total distance travelled by the robot per second divided by the total object moving distance per second. The rationale is that if the participants share more common ground, the collaborators would have a better expectation of the others' actions, thereby allowing better action coordination so that the task execution would become more efficient. More efficient performance will result in the reduction of task completion time and less vehicle movement.

In Japanese, the demonstratives "*kore*," "*sore*," and "*are*" play the same role as the English pronoun "it" for objects in addition to their traditional deictic expressions. For this reason, we have categorized these words simply as demonstratives.

2 Method

2.1 Participants and Construction of Corpus

The participants were all Japanese undergraduate volunteers whose first language was Japanese. Participants had some experience with computer games and joy-sticks. There were four groups of three participants (a total of 12 participants; 12 males; age range = 21–24; M age = 22.3; SD = 0.98), and they performed a virtual collaborative conveyer task. The corpus data were transcribed by two trained graduates and two trained undergraduates, using the corpus constructing analysis software CLAN [22]. The transcription format follows the Japanese *Wakachi* format for Japanese utterances [23,24].

2.1.1 Task

Figure 1 illustrates the experimental setup. The collaborative task was to transfer objects to the designated positions by robot vehicles in a virtual space. Three participants manipulated each robot vehicle with a joystick input device. There were three objects, and each object was colored green, blue, and red, respectively. The whole task was designed so that collaboration among the robot vehicles is needed to complete the task. For example, each vehicle was relatively small compared to the larger and heavier load boxes. The task must be completed within the quickest time and with the least crashes. Every vehicle crash against the wall is a penalty for its operator, and every object crash is a penalty for the group. The participants were informed that payment to them would be calculated according to the completion time and number of

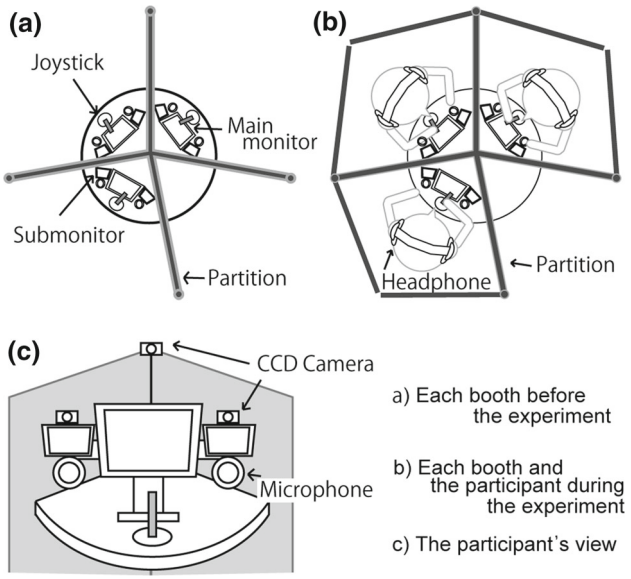

Fig. 1 The experimental setup

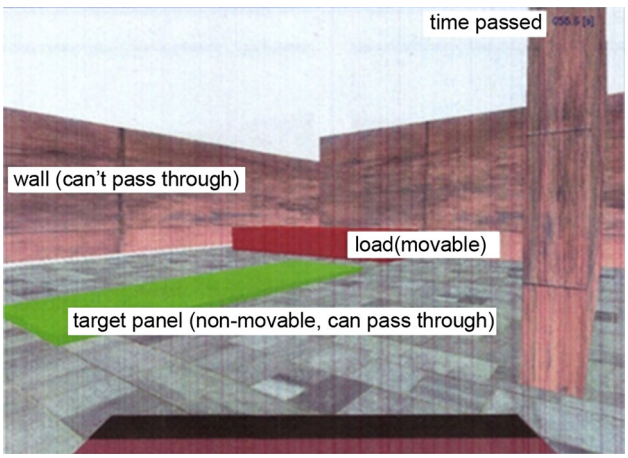

Fig. 2 The participant's view of the virtual task field. The front part of the participant's robot vehicle is shown at the bottom of the monitor

penalties. Each booth, where a participant was seated, had been separated by a wooden partition to disable eye communication. However, to complete the task, the participants were encouraged to talk and collaborate through microphones, headphones, and CCD cameras. Each group was asked to complete the same task ten times. Figure 2 shows an example of a virtual task field through the point-of-view shot as shown on the main display screen.

Figure 3 shows the position of each vehicle, the load box, and the target panel. The field was separated by walls. Figure 4 shows the aerial view of the task field and the typical movements of the boxes. All the load boxes had to pass through relatively narrow openings between the walls with the consideration of enabling more participant collaboration. There was no route restriction. However, this aerial view is not available to participants in the experiment.

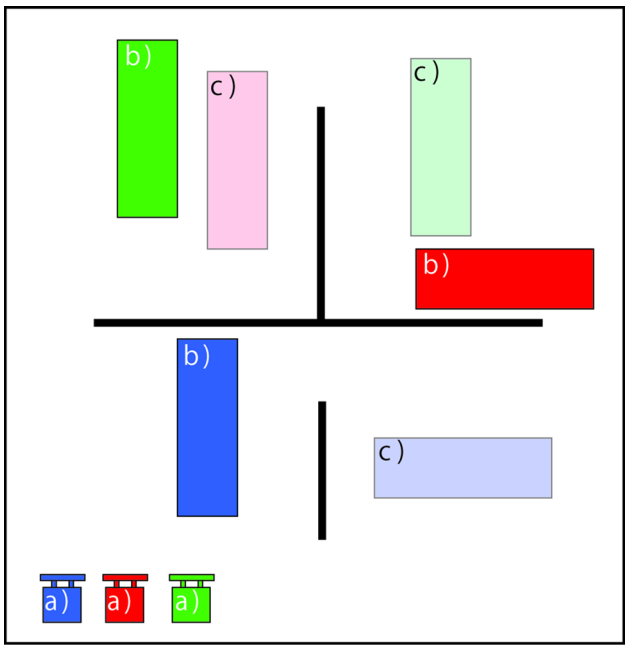

Fig. 3 Locations of robot vehicles (**a**), load boxes (**b**), and target panels (**c**). Load boxes were to be transported to each corresponded target panels by robot-vehicles

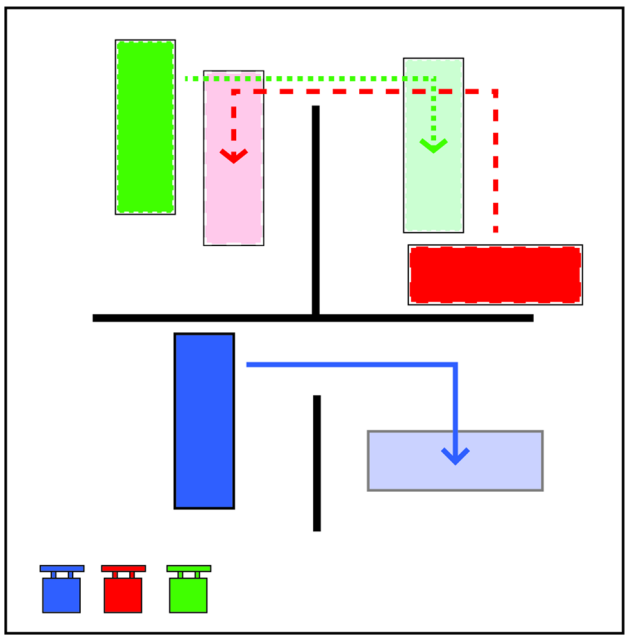

Fig. 4 An example of each transportation route in the aerial view of the task field

Participants would look at the virtual experimental field on the main display panel to manipulate their vehicles. In the sub-displays placed on both sides of the main panel, the faces of the two participants in the group were shown. For this reason, participants could check their team members' involvement through their facial expressions and mouth

movements. A joystick in front of each participant could be manipulated to move the robot forward and backward, and the device can perform turning movements. The log data of the three vehicles' and objects' positions in the time course were obtained.

2.2 Analysis of Skill Development

We measured skill development using task completion time and the travelling distance of the robots and objects. The travelling distance of robots d^r and objects d^o, with a sampling count of task completion N, is defined as follows:

$$d^r = \sum_{k=1}^{N} \sum_{i=1}^{3} \frac{|v_i^r[k]|}{3N},$$

$$d^o = \sum_{k=1}^{N} \sum_{i=1}^{3} \frac{|v_i^o[k]|}{3N},$$

where $v_i^r[k]$ and $v_i^o[k]$ denote the velocity of the robots and objects at the sampling count k, respectively. Their ratio, d^r/d^o, is utilized for the group performance index. The experts would convey the objects smoothly and optimize the robot control with the least motion for the task completion. As a result, the d^r/d^o will gradually decrease as the skill develops.

2.3 Analysis of the Corpus Data

We conducted a morphological analysis on the corpus data for each trial, using the MOR and POST program of CLAN. The FREQ program automatically lists all the morphemes and frequencies of the target corpus. The word class categories we computed in this study were demonstratives and nouns, based on the MOR program. The demonstratives included the following: (a) "*kore*" (this), "*koko*" (here), and other *ko*-category demonstratives, (b) "*sore*" (that-proximal), "*soko*" (there), and other *so*-category demonstratives, and (c) "*are*" (that-distal), "*asoko*" (there), and other *a*-category demonstratives. Concerning nouns, we further classified the nouns into four types, based on the context, not by the software but by a trained annotator. The four types were a) common nouns (e.g., car, box), b) direction nouns (e.g., left, straight), c) place nouns (e.g., chink, aisle), and d) time nouns (e.g., next, after).

3 Results

3.1 Skill Development

Figure 5 shows the mean task completion time in each trial of the four groups. We examined the skill development using a correlation analysis of all groups. The result showed a significant negative correlation between the task completion time and the number of trials ($r = -0.815$, $p = 0.004$). The task completion time decreased in all groups when they repeated the task.

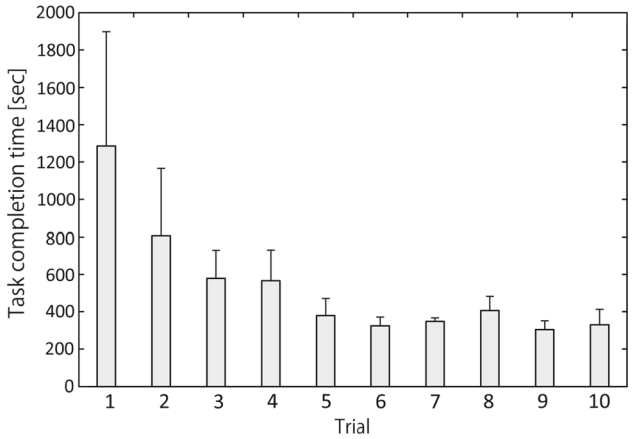

Fig. 5 Task completion time in each trial. The error bars denote the standard deviation

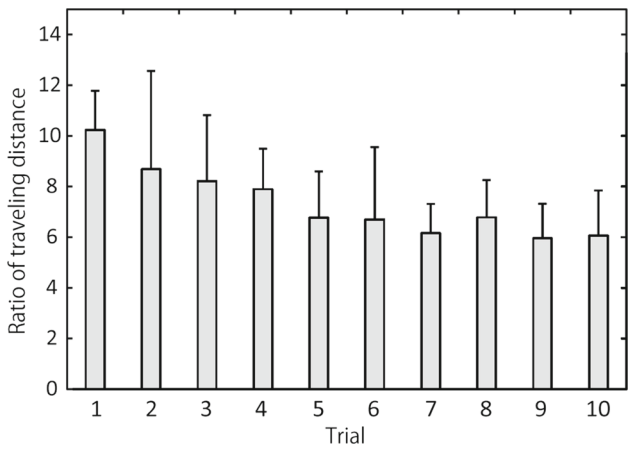

Fig. 6 The ratio of the robot travelled distance and the object travelled distance in each trial. The error bars denote the standard deviations

The result of the efficiency measure of the robots (Fig. 6) showed that all groups successfully finished the task in all 10 trials. The mean distance the robots traveled strongly and negatively correlated with the number of trials ($r = -0.916$, $p < 0.001$). They generally moved more effectively to complete the task, thereby suggesting that the participants increased the efficiency of their robot movements.

3.2 The Use of Nouns and Demonstratives

Figure 7 shows the use of demonstratives and nouns. To examine whether reduced use occurred, we computed the correlation between the mean frequency of demonstratives

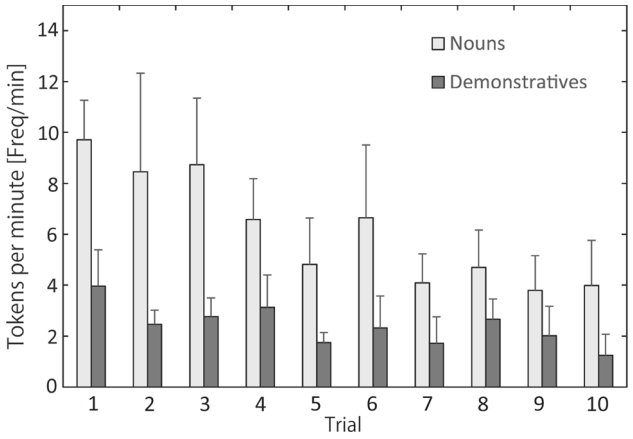

Fig. 7 The use of nouns and demonstratives in each trial. The error bars denote the standard deviation

Table 1 Pearson's correlation between the mean frequency in each type of nouns and the number of trials

	Correlation coefficient	p value	Sig.
Common	− 0.845	0.002	**
Direction	− 0.850	0.002	**
Place	− 0.865	0.001	**
Time	− 0.379	0.279	n.s.

$^{**}p < 0.01$; $^{*}p < 0.05$

and nouns, and the number of trials. The frequency of nouns per minute negatively correlated with the number of trials ($r = -0.918$, $p < 0.001$). The frequency of demonstratives per minute also negatively correlated with the number of trials ($r = -0.740$, $p = 0.014$).

3.3 Use of Different Types of Nouns

Figure 8 shows the use of each noun type in each trial. To examine whether reduced use occurred, we computed the correlation between the mean frequency of each type of noun and the number of trials (Table 1). The correlation between the mean common noun per minute and the number of trials was negatively significant ($r = -0.845$, $p = 0.002$). The correlation between the mean direction noun per minute and the number of trials was negatively significant ($r = -0.850$, $p = 0.001$). The correlation between the mean space noun per minute and the number of trials was negatively significant ($r = -0.865$, $p = 0.001$). However, the correlation between the mean time noun per minute and the number of trials was not significant ($r = -0.379$, $p = 0.279$, n.s.). Because the frequency of time nouns was very low throughout the trials, this might be caused by a floor effect.

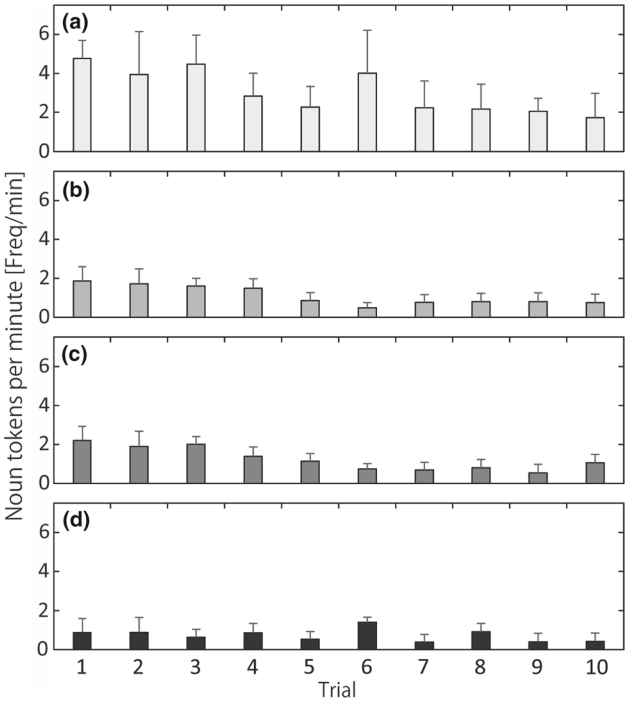

Fig. 8 The use of each noun type in each trial. The error bars denote the standard deviation. The row with **a** denotes common nouns, **b** direction nouns, **c** space nouns, and **d** time nouns

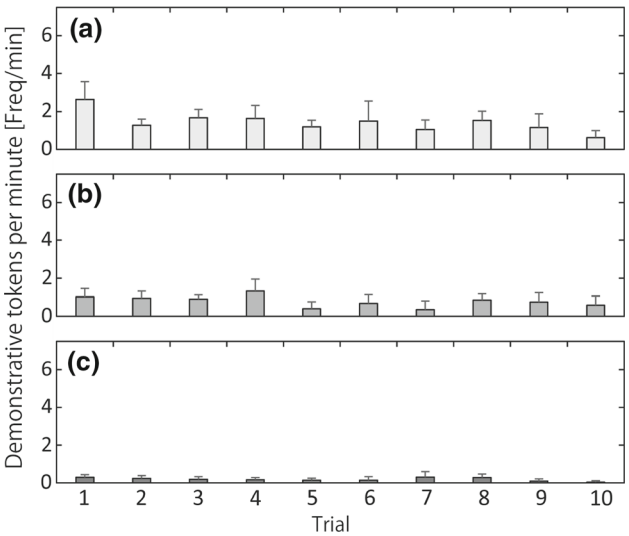

Fig. 9 The mean use of each demonstrative type in each trial in the four groups. The error bars denote the standard deviation. The row with **a** denotes the demonstrative "This," **b** denotes the "That-Proximal," and **c** denotes the "That-Distal"

3.4 The Use of Different Types of Demonstratives

Figure 9 shows the mean frequency of each demonstrative per minute in each trial in the four groups. To examine whether reduced use occurred, we computed the correlation between the mean frequency in each type of demonstrative

Table 2 Pearson's correlation between the mean frequency in each type of demonstratives and the number of trials

	Correlation coefficient	p value	Sig.
This	− 0.730	0.016	*
That-proximal	− 0.507	0.135	n.s.
That-distal	− 0.504	0.138	n.s.

$^{**}p < 0.01$; $^{*}p < 0.05$

and the number of trials (Table 2). The correlation between the mean frequency of the demonstrative "This" type per minute and the number of trials was negatively significant ($r = -0.730$, $p = 0.016$). However, the demonstratives "That-Proximal" and "That-Distal" did not correlate with the number of trials (That-Proximal: $r = -0.507$, $p = 0.135$, n.s.; That-Distal: $r = -0.504$, $p = 0.138$, n.s.). Because the frequency of these demonstrative types was relatively low throughout the trials, this might be caused by a floor effect.

4 Discussion

The task completion time decreased. This suggests that the participants' skill to complete the task developed as they repeated the task. The ratio of robot movement over object movement also decreased. This demonstrates that the robots moved less with increased efficiency. These data suggest the possibility of the establishment of a common ground among the participants during the repetition of the task. Here, we must point out that task proficiency itself may also contribute to the reduction. However, to a certain extent, we can say that knowledge of one's own and others' changing proficiency states may also be included in the grounding process.

The results indicate that the use of nouns and demonstratives per minute decreased as the participants repeated the task. The speakers talked using fewer referential words. This occurred in both symbolic references (nouns) and deictic references (demonstratives). This tendency continued as the speakers accumulated the experience of joint action. Thus, the result confirmed the previous finding that the word use for referents was reduced in repeated referencing [3]. We further added new evidence to the literature, in that the use of both symbolic and deictic references decreases in repeated referencing and in a similar degree in these two types. We also added that the reduction of referential words occurs in a joint action of three people.

Amongst the demonstratives, the use of the "This" type decreased as the task was repeated with the grounding process. The "This" demonstrative is also the most frequently used in the task. The participants favored the use of "This" to attend jointly to the same referent during the task. Unlike the "This" demonstrative, the "That-proximal" were not frequently used, and "That-distal" were only rarely used. This phenomenon might have been possible because the vehicles were typically close to each other in collaborative settings. In addition, the participants might have felt that the virtual space presented in a computer monitor was relatively small.

The study shows that human–robot collaboration designs must deal with reduction of referential expressions as humans establish a common ground. Why does the reduction of reference expressions occur in human–human interactions? The referents in a given task situation would become increasingly predictable in a repeated human–robot joint action. Communication with fewer words means more shared knowledge between the collaborators. Typically, people do not have to use nouns when the referents are obvious, with or without referencing or using other reference words, such as demonstratives. In addition, people do not even need demonstratives because the referent is already shared in the on-going grounding process. In fact, this kind of discourse interaction may be perceived as more "natural" and "human-like" by human collaborators. The reason is that Grice's Cooperative Principle [25] on the quantity of information transmission is satisfied [7]. In addition, fewer resources may be allocated in working memory [26] for information exchange on the repeated referring. The study strongly suggests that a proper design for a human–robot joint action should consider the process of grounding and reduction of referring. Then how do social robots adapt themselves for reduction of referring in human–human collaboration? This question is beyond the scope of this study, but our study does suggest that robots must have some means by which to estimate relevant referents. Assessing the exact state of common ground among collaborators is necessarily very important.

5 Conclusion

In this study, we examined the language use of referring in conversations when people jointly act on objects in a collaborative task. We extracted nouns and demonstratives as linguistic forms for referring. We then examined whether the use of these words changes when people repeatedly executed the same collaborative task. We used a virtual space for the collaborative task because we wanted to control the common ground that each group of participants shared. The study provided clear evidence that the reduction in the use of referential words occurred in both nouns and demonstratives. This phenomenon seemed to occur with the grounding process. The study suggests that an appropriate design of human–robot collaboration must account for grounding and reduction of referencing.

The use of a common ground occurs in various types of knowledge and in many layers at each moment of joint action

[7]. The present study did not analyze specifically how such a reduction of referential words is achieved with grounding. The nature of efficient language use and effortless information transmission must be explored by considering such layers. We are also aware that there is still a need to examine whether task proficiency itself is a factor. Also, it can be said that knowledge of one's own and others' changing proficiency states may be included in the grounding process. Future research that requires the robot to perform few target objectives, such as moving the box to panel A, B, and C in this predetermined order may show the relationship between a certain reduction of word use and the efficiency of robot movement.

Acknowledgements We thank all the volunteers who have participated in this study. We also thank the reviewers for their insightful comments and suggestions on the manuscript. This work was supported by MEXT/JSPS KAKENHI Grant Number JP17H06382 in #4903 (Evolinguistics), JP16K04318, JP 15K06153, JP15K05912, and JP26870549.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Tomasello M (2009) Why we cooperate. MIT Press, Cambridge
2. Sebanz N, Bekkering H, Knoblich G (2006) Joint action: bodies and minds moving together. Trends Cogn Sci 10:70–76
3. Brennan SE, Galati A, Kuhlen AK (2010) Two minds, one dialog: coordinating speaking and understanding. Psychol Learn Motiv 53:301–344
4. Garrod S, Pickering MJ (2004) Why is conversation so easy? Trends Cogn Sci 8:8–11
5. Schober MF, Brennan SE (2003) Processes of interactive spoken discourse: the role of the partner. In: Graesser AC, Gernsbacher MA, Goldman SR (eds) Handbook of discourse processes. Lawrence Erlbaum Associates, Mahwah, pp 123–164
6. Heim I (1982) The semantics of definite and indefinite noun phrases. PhD Thesis, University of Massachussets, Amherst
7. Clark HH (1996) Using language. Cambridge University Press, Cambridge
8. Clark HH, Schaefer EF (1989) Contributing to discourse. Cogn Sci 13:259–294
9. Clark HH, Schreuder R, Buttrick S (1983) Common ground at the understanding of demonstrative reference. J Verbal Learn Verbal Behav 22:245–258
10. Kobayashi H, Yasuda T, Igarashi H, Suzuki S (2012) Changes of action ontology in conversation among collaborators using virtual space. In Proceedings of 21st IEEE international symposium on robot and human interactive communication, pp 748–752
11. Ballard DH, Hayhoe MM, Pook PK, Rao RP (1997) Deictic codes for the embodiment of cognition. Behav Brain Sci 20:723–742
12. Kooijmans T, Kanda T, Bartneck C, Ishiguro H, Hagita N (2007) Accelerating robot development through integral analysis of human–robot interaction. IEEE Trans Robot 23:1001–1012
13. Sato E, Yamaguchi T, Harashima F (2007) Natural interface using pointing behavior for human–robot gestural interaction. IEEE Trans Ind Electron 54:1105–1112
14. Sugiyama O, Kanda T, Imai M, Ishiguro H, Hagita N (2007) Natural deictic communication with humanoid robots. In: Proceedings of the 2007 IEEE/RSJ international conference on intelligent robots and systems, pp 1441–1448
15. Gentner D (1982) Why nouns are learned before verbs: linguistic relativity versus natural partitioning. In: Kuczaj S (ed) Language development: language, cognition, and culture. Lawrence Erlbaum Associates, Hillsdale, pp 301–334
16. Diessel H (1999) Demonstratives: form, function and grammaticalization. John Benjamins, Philadelphia
17. Diessel H (2006) Demonstratives, joint attention, and the emergence of grammar. Cogn Ling 17:463–489
18. Coventry KR, Valdés B, Castillo A, Guijarro-Fuentes P (2008) Language within your reach: near–far perceptual space and spatial demonstratives. Cognition 108:889–895
19. Coventry KR, Griffiths D, Hamilton CJ (2014) Spatial demonstratives and perceptual space: describing and remembering object location. Cogn Psychol 69:46–70
20. Takahashi T, Suzuki M (1982) Referent area of demonstratives KO/SO/A (Ko So A no shizi ryoiki ni tsuite). Report of Nati Insti Jap Lan Ling 71:1–44 **(in Japanese)**
21. Endo M (1988) Influence of operability and the dehumanization of the hearer on the use of three sets of Japanese demonstratives: Ko So A. Shinrigaku Kenkyu 59:199–205 **(in Japanese with English abstract)**
22. MacWhinney B (2000) The CHILDES project: tools for analyzing talk, 3rd edn. Lawrence Erlbaum Associates, Mahwah
23. Miyata S, Morita H, Muraki K (2004) Using database of utterances from today: first-step for using child language data exchange system (Kyo kara tsukaeru hatsuwa deetabeesu: shoshinsha no tame no CHILDES nyumon). Hitsuji-Shobo **(in Japanese)**
24. Oshima-Takane Y, MacWhinney B, Sirai H, Miyata S, Naka N (1998) CHILDES for Japanese, 2nd edn. The JCHAT Project, Chukyo University, Nagoya
25. Grice HP (1975) Logic and conversation. In: Cole P, Morgan J (eds) Studies in syntax and semantics III: speech acts. Academic Press, New York, pp 183–198
26. Baddeley AD (2007) Working memory, thought and action. Oxford University Press, Oxford

Harumi Kobayashi received her M.A. and Ph.D. degrees in Developmental Psychology at the University of Maryland, U.S., in 1985 and 1991, respectively. She is a professor in the Department of Information System Design at Tokyo Denki University. She is also a licensed Clinical Developmental Psychologist. Her major research interests are psycholinguistics, cognitive science, and language development. She has been a current president of the JSLS (Japanese Society for Language Sciences) since 2013. She is also a member of the JCSS (Japanese Cognitive Science Society) and JSDP (Japanese Society for Developmental Psychology).

Tetsuya Yasuda received his B.S. and M.S. degrees and his Doctor of Informatics degree at Tokyo Denki University, Saitama, Japan, in 2004, 2006, and 2011, respectively. He was a Research Associate in the Department of Human Developmental Psychology at Jumonji University during this research project. He is currently a designated

assistant professor at Tokyo Denki University. His current research interests include human–human interaction, ostensive communication, and pragmatics interpretations. He is a member of the JCSS (Japanese Cognitive Science Society), JSLS (Japanese Society for Language Sciences), and JSDP (Japanese Society for Developmental Psychology).

Hiroshi Igarashi received his B.S., M.S., and Ph.D. degrees from the Department of Electrical Engineering at Tokyo Denki University in 2000, 2002, and 2005, respectively. He is currently an associate professor in the Department of Electrical and Electronic Engineering at Tokyo Denki University. His major research interests are robotics, human–machine systems, and artificial intelligence. He is a member of the IEEE, IEEJ (Institute of Energy Economics, Japan), RSJ (Robotics Society of Japan), and JSME (Japan Society of Mechanical Engineers).

Satoshi Suzuki received his B.S. degree in Control Engineering, M.S. degree in the Department of Systems Science, and Ph.D. in the Department of Mechanical and Control Engineering from Tokyo Institute of Technology in 1993, 1995, and 2004, respectively. He is currently an associate professor in the Department of Robotics and Mechatronics at Tokyo Denki University. His major research interests are human–machine systems, bioinstrumentation, robotics, and service engineering. He is a member of the IEEE, SICE (Society of Instrument and Control Engineers), RSJ (Robotics Society of Japan), IEEJ (Institute of Energy Economics, Japan), and Society for Serviceology.