



OPEN ACCESS

EDITED BY

Renia Lopez-Ozieblo,
Hong Kong Polytechnic University,
Hong Kong SAR, China

REVIEWED BY

Dimitra Anastasiou,
Luxembourg Institute of Science and
Technology, Luxembourg
Lisa-Marie Krause,
Julius Maximilian University of Würzburg,
Germany

*CORRESPONDENCE

Kazuki Sekine
✉ ksekine@waseda.jp

RECEIVED 02 May 2025

ACCEPTED 29 September 2025

PUBLISHED 23 October 2025

CITATION

Sekine K, Kanemaru K and Kadota K (2025)
The facilitative role of listener's pointing
gestures
in collaborative tasks.
Front. Commun. 10:1621867.
doi: 10.3389/fcomm.2025.1621867

COPYRIGHT

© 2025 Sekine, Kanemaru and Kadota. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

The facilitative role of listener's pointing gestures in collaborative tasks

Kazuki Sekine^{1*}, Kotaro Kanemaru² and Keisuke Kadota¹

¹Faculty of Human Sciences, Waseda University, Tokorozawa, Japan, ²School of Human Sciences, Waseda University, Tokorozawa, Japan

We investigated the role of listener-produced pointing gestures in a collaborative sticker localisation task. While previous research has emphasised the communicative value of speaker gestures, few experimental studies have examined how listeners' gestures shape interaction. We examined how listeners' gestures facilitate interaction, using a two-by-two within-subjects design that manipulated whether speakers were allowed to gesture and whether listeners began trials with a pointing gesture. Forty-eight adults participated in a sticker localisation task, and three dependent measures were analysed: task completion time, the number of spatial utterances, and gesture duration. The results demonstrated that listeners' pointing gestures significantly reduced task duration, regardless of whether the speaker gestured. Additionally, these listener gestures prompted longer gestural output by the speakers, suggesting that visible bodily engagement from listeners influenced speakers' multimodal behaviour. By contrast, speaker gestures did not significantly affect efficiency. These findings provide empirical support for the idea that listeners' gestures function as participatory and epistemic actions, not merely as passive cues of understanding. The study supports a reciprocal model of gesture, demonstrating that both speakers and listeners use bodily actions to co-construct spatial reference. By providing experimental evidence on listener gestures, it contributes to research that frames gesture as an interactive and embodied process. These findings also suggest potential applications for designing collaborative systems that respond to real-time bodily cues.

KEYWORDS

listener's pointing gestures, interaction, collaborative work, common ground, multimodal communication

1 Introduction

People often produce gestures when they speak, known as co-speech gestures. During face-to-face communication, these gestures accompany speech and serve as essential multimodal resources for coordinating shared attention and supporting joint action. Amongst these, pointing gestures are especially important in spatial tasks, as they help clarify referents and reduce verbal ambiguity. Previous research has shown that speakers' gestures contribute to the efficiency of collaborative activity by grounding verbal instructions in the physical environment (McNeill, 1992; Kendon, 2004). For instance, Kang and Tversky (2016) demonstrated that pointing gestures by speakers improved listener comprehension during instructional tasks. Similarly, Bentley et al. (2023) found that the use of iconic gestures depicting the shape or action of the referent enhanced students' understanding in classroom settings. These findings suggest that gestures not only supplement verbal content but also function as integral components of interactive meaning-making.

While most studies have focused on speaker-produced gestures, recent work has begun to highlight the communicative role of gestures produced by listeners. These gestures are not merely reactive or peripheral but can play an active role in shaping the flow of interaction. [Holler and Wilkin \(2011\)](#) demonstrated that, in later conversational turns, listeners sometimes mimic a speaker's gesture form, thereby signalling alignment and mutual understanding. In addition, [Healey et al. \(2015\)](#) observed that listeners used gestures during clarification sequences, often to express their interpretation of the speaker's message or to offer alternative understandings. Similar findings have also been reported in other studies (e.g., [Kimbara, 2006](#); [Holler et al., 2018](#); [Sekine and Özyürek, 2024](#)). These studies suggest that listener gestures may function as epistemic and participatory actions that support collaborative communication. However, most of these findings are based on observational data, with few studies having experimentally tested the effect of listener gestures on measurable outcomes such as task efficiency or speaker behaviour.

A notable example is provided by [Hosoma et al. \(2004\)](#), who examined a spatial localisation task in which one participant served as an instructor and the other as a searcher. When the searcher pointed to a guessed location on her helmet, the instructor used this gesture as a reference point for subsequent instructions, leading to successful task completion. While this study offered valuable insights into how listener gestures can serve as resources for the speaker, it was qualitative in nature and lacked experimental validation. Thus, it remains unclear whether such effects generalise to broader collaborative contexts.

To address this gap, the present study investigates how listener-produced pointing gestures influence task performance and speaker behaviour in a controlled face-to-face spatial task. Building on earlier research on speaker gestures ([Kang and Tversky, 2016](#); [Bentley et al., 2023](#)) and listener gestures ([Hosoma et al., 2004](#); [Holler and Wilkin, 2011](#)), we examine whether gestures by listeners facilitate the construction of shared spatial reference and affect the communicative strategies employed by speakers. Using a two-by-two within-subjects design, we manipulated whether the instructor was allowed to gesture and whether the searcher began the trial with a pointing gesture in place. We then measured three dependent variables: task completion time, the number of spatial utterances, and the total duration of instructor gestures. Both speaker and listener gestures were recorded to allow a symmetrical analysis of their contributions to joint activity. While task completion time served as our primary index of efficiency, we also included gesture duration as a complementary measure. This allowed us to examine whether improvements in collaborative performance arose from more economical use of gestures, more efficient verbal communication, or from their interplay. Considering gesture duration alongside utterance counts also enabled us to capture how communicative behaviours were organised and adapted across conditions.

We tested two hypotheses. First, we predicted that speaker gestures would reduce task duration by clarifying spatial referents, reducing verbal ambiguity, and anchoring instructions in the physical space, thereby increasing the clarity and effectiveness of spatial instructions, consistent with previous findings on gesture-assisted communication ([Kang and Tversky, 2016](#); [Bentley et al., 2023](#)). Second, we hypothesised that listener gestures, particularly those visible from the start of the task, would serve as anchoring cues that speakers could adapt to, thereby enhancing efficiency. This second hypothesis was motivated by prior work ([Hosoma et al., 2004](#); [Holler and Wilkin, 2011](#)) showing that listener gestures can serve as visible

indicators of comprehension and can shape how speakers formulate their instructions. In addition, the measures of spoken utterances and gesture duration were included to examine how listener gestures influenced instructors' verbal and non-verbal behaviour. These analyses were partly exploratory but grounded in the expectation that listener gestures function as interactive components of the exchange.

2 Method

2.1 Experimental design

This study employed a two-by-two within-subjects factorial design to investigate how gestures produced by both the speaker (instructor) and the listener (searcher) influence performance in a collaborative spatial task. The two independent variables were: the presence or absence of pointing gestures by the instructor, and the initial placement of the searcher's index finger on the board. The conditions for each independent variable are outlined below.

2.1.1 Instructor's pointing

- Pointing condition: The instructor was allowed to freely use pointing gestures to indicate the location of a sticker and was encouraged to gesture as much as possible during the trial.
- No pointing condition: the instructor was instructed to refrain from using any gestures and was required to keep both hands behind their back throughout the trial.

2.1.2 Searcher's pointing placement

- Pointing placement condition: The searcher began each trial with their index finger placed at the centre of a plastic board (see [Figures 1, 2](#)) and was asked to keep their finger raised throughout the trial.
- No pointing placement condition: The searcher began with their hands on their lap and could only raise their finger after receiving verbal instructions from the instructor. After successfully identifying a sticker, the searcher returned their hand to their lap before proceeding to the next target. This condition restricted spontaneous gesturing at the beginning of each trial but allowed guided pointing during the search phase.

To minimise time loss during finger lowering, instructors were instructed to continue giving guidance even while the searcher was returning their hand to their lap. The four resulting experimental conditions are summarised in [Table 1](#).

2.2 Planned analysis

The design was a 2 (Instructor's Pointing: present vs. absent) $\times$ 2 (Searcher's Pointing Placement: present vs. absent) repeated-measures design. Because each dyad experienced all four conditions, the dyad served as the unit of analysis. This choice follows previous research on gesture and collaborative tasks that have similarly analysed dyads using repeated-measures ANOVA (e.g., [Bangerter and Chevalley,](#)



FIGURE 1
Start of the task in the searcher's pointing placement condition.



FIGURE 2
A searcher with a plastic board and four coloured stickers.

2007; Kraut et al., 2003; Wang et al., 2021). Treating the dyad as the unit also helped to reduce the risk of practice or learning effects that could arise if each individual participant experienced all conditions in sequence. Conceptually, the data could be modelled as participants nested within dyads, which would require a more complex random-effects structure. To avoid this complexity, we applied a repeated-measures ANOVA, which only assumes sphericity. We also ran exploratory Linear Mixed Models (LMMs) including nested random effects of participants within dyads. These analyses showed signs of overdispersion, but importantly, the statistical significance of the fixed effects did not differ from those obtained with ANOVA. We therefore

TABLE 1 Usage of pointing gestures by the instructor (I) and the searcher (S) across the four experimental conditions.

Condition	Instructor's pointing	Searcher's pointing placement
I pointing—S pointing	✓	✓
I no pointing—S pointing	–	✓
I pointing—S no pointing	✓	–
I no pointing—S no pointing	–	–

I, instructor; S, searcher.

report the ANOVA results for consistency and comparability with prior literature. We further checked the assumptions of the ANOVA. Shapiro–Wilk tests indicated that not all conditions strictly met the assumption of normality; however, ANOVA is generally robust to moderate departures from normality in within-subject designs. Because our design only included two-level factors, the assumption of sphericity is automatically satisfied and was not separately tested.

2.3 Participants

Forty-eight native Japanese-speaking adults (24 males, 22 females, two unspecified; age range = 18–26 years, $M_{age} = 20.32$, $SD = 2.30$) participated in the study. All participants were recruited from a university student population and had no prior familiarity with each other. They were randomly paired into 24 dyads. Each dyad completed all four conditions in a counterbalanced order to control for sequence effects. Roles (instructor or searcher) were switched after 12 trials so

that all participants experienced both roles. This study was approved by the Ethics Committee of Waseda University (Approval No. 2023–125), and all participants gave written informed consent prior to participation.

2.4 Materials

A plastic board (150 mm × 200 mm × 1 mm), four stickers (diameter = 4 mm; red, blue, yellow, and black), and rubber bands were used to create the apparatus (Figure 2). The board was divided into four quadrants by invisible vertical and horizontal lines intersecting at the centre, with one sticker placed randomly within each quadrant. A total of 24 boards were prepared in advance for 24 trials.

2.5 Procedures

The task for participants involved a sticker localisation game, in which the instructor gave verbal instructions to help the listener identify the positions of stickers attached to a board worn on the listener's forehead. Each experimental session was conducted with a single dyad consisting of an instructor and a searcher. At the beginning of the session, participants were informed about the rules for each condition (i.e., whether pointing gestures were permitted or not). They were reminded of the relevant rule again immediately before each block began to ensure compliance. Both the instructor and the searcher were allowed to speak freely during the task, with no constraints on what they could say or when.

In each trial, the participant in the listener role wore a transparent plastic board (150 mm × 200 mm × 1 mm) fixed to the front of the head with a rubber band, such that they could not see the stickers but were able to point at them. The speaker sat 2.5 m in front of the listener and had a full view of the board and the listener's gestures. The pair sat on chairs, and the interaction was recorded with two cameras: one placed in front of the listener (as in Figure 1) and another positioned diagonally from the side to capture both participants' gestures.

At the start of each trial, four coloured stickers (red, blue, yellow, and black; diameter = 4 mm) were randomly affixed to the front of the board, with one sticker randomly placed within each quadrant. The listener was responsible for identifying their positions using their index finger, while the instructor provided verbal instructions to help locate them. The listener was instructed to find the stickers in a fixed order (red, blue, yellow, black). A trial ended when both participants mutually agreed that all four stickers had been located. Each dyad completed 24 trials in total (six per condition), with conditions presented in randomised order. After 12 trials, participants switched roles so that each served as both instructor and listener. Each trial typically lasted around 50–60 s, and a complete session of 24 trials took approximately 25–30 min. Short breaks were provided between blocks, and no participants reported fatigue or discomfort.

To maintain consistency and eliminate confounding factors, three task constraints were imposed. First, neither participant was allowed to stand or physically touch the plastic board, and the instructor was explicitly prohibited from pointing directly at the stickers on the board. However, the instructor was permitted to use pointing gestures

directed towards her own body. This constraint was implemented to prevent the task from becoming trivially easy. Second, the searcher was instructed not to rub or slide her finger across the board surface while searching, so that tactile cues could not be used to locate stickers. Third, participants were encouraged to complete the task as quickly as possible. This time pressure was introduced to capture the efficiency of collaborative performance and reduce unnecessary hesitation or redundant actions, ensuring that performance differences reflected communicative efficiency rather than deliberate pacing.

2.6 Dependent measures

Three dependent variables were analysed for each trial: task completion time, the number of spatially instructive utterances produced by the instructor, and the duration of the instructor's pointing gestures. The primary measure was task completion time, defined as the total duration from the experimenter's cue to begin the task to the moment the searcher successfully located the fourth sticker, which served as an index of collaborative efficiency.

The second variable was the number of spatially instructive utterances produced by the instructor. Utterances were coded at the clause level when they directly referred to the sticker's location, using spatial terms (e.g., “top left,” “move 2 cm right”), directional guidance (e.g., “keep going,” “you went too far”), or demonstratives accompanied by gesture (e.g., “this way”). This measure served as an index of the instructor's verbal effort during the task. All speech data were initially annotated by the second author, and the first author independently re-annotated 25% of the dataset. Inter-rater agreement was high (Cohen's $\kappa = 0.94$), and any discrepancies were resolved through discussion.

The third variable was the duration of the instructor's pointing gestures. Gesture duration was measured based on the gesture phase framework established in prior research (McNeill, 1992). Each gesture was segmented into four phases: preparation, stroke, post-stroke hold, and retraction. Duration was defined as the total time from the onset of the preparation phase to the end of the retraction phase. When multiple gestures occurred within a single trial, their durations were summed to calculate the total gesture duration for that trial. All gesture data were annotated using ELAN software (Lausberg and Sloetjes, 2009), which enabled frame-by-frame analysis of hand movements. Two trained coders independently annotated 25% of the dataset, and inter-rater reliability was assessed using Cohen's kappa ($\kappa = 0.86$), indicating high agreement. Any discrepancies were resolved through discussion. This method allowed for a precise distinction between trials involving multiple brief gestures and those with a single prolonged hold, thus addressing concerns about overgeneralised gesture metrics in prior literature.

3 Results

3.1 The influence of pointing gestures on the task completion time

To examine how gestures influenced collaborative efficiency, we analysed task completion time (in seconds) across the four experimental conditions. For each dyad, we first calculated the average

task completion time over six trials per condition. Then, we computed the overall mean of these averages across all dyads as follows. The mean completion time in the Searcher Pointing Placement—Instructor Pointing condition was 51.49 s (SD = 14.61), compared to 55.73 s (SD = 14.27) in the Searcher Pointing Placement—Instructor No Pointing condition. For the Searcher No Pointing Placement—Instructor Pointing condition, the mean time was 59.95 s (SD = 17.76), and for Searcher No Pointing Placement—Instructor No Pointing, it was 60.63 s (SD = 13.78). These values are illustrated in Figure 3.

A two-way repeated-measures ANOVA was conducted with instructor's pointing and searcher's pointing placement (i.e., whether the listener began the trial with their index finger placed at the centre of the board or with their hands on their lap) as within-subject factors and task completion time as the dependent variable. The analysis revealed no significant interaction between the two factors, $F(1, 23) = 0.94$, $p = 0.34$, partial $\eta^2 = 0.04$. There was also no main effect of instructor's pointing, $F(1, 23) = 1.07$, $p = 0.31$, partial $\eta^2 = 0.04$. However, we observed a significant main effect of searcher's pointing placement, $F(1, 23) = 6.14$, $p = 0.021$, partial $\eta^2 = 0.21$. These results suggest that the visibility of searcher's initial pointing placement facilitated faster task completion, regardless of whether the instructor gestured.

3.2 Number of utterances of instructions indicating the location of the sticker

To examine whether the reduced task completion time observed in the pointing placement condition was associated with differences in instructors' communicative behaviour, we analysed the number of spatially instructive utterances produced by the instructor across conditions. This analysis was included in our design to evaluate whether efficiency gains were reflected in instructors' verbal effort as well as gestural behaviour. Because reduced task completion time might reflect more targeted and efficient communicative behaviour by the instructor, we also examined whether there were differences in instructors' gestures across conditions.

For each dyad, the mean and standard deviation were calculated for the number of utterances used to indicate the position of the sticker across six trials per condition. We then computed the overall mean of these values across all dyads. The mean number of instructive utterances was 13.98 (SD = 4.10) in the Searcher Pointing

Placement—Instructor Pointing condition, 17.46 (SD = 3.14) in the Searcher Pointing Placement—Instructor No Pointing condition, 13.97 (SD = 7.43) in the Searcher No Pointing Placement—Instructor Pointing condition, and 16.94 (SD = 3.10) in the Searcher No Pointing Placement—Instructor No Pointing condition (see Figure 4).

A two-way ANOVA with the same independent variables showed no significant interaction between searcher and instructor conditions, $F(1, 23) = 2.21$, $p = 0.15$, partial $\eta^2 = 0.09$. Similarly, no main effects were found for either instructor's pointing, $F(1, 23) = 0.12$, $p = 0.73$, partial $\eta^2 = 0.01$, or searcher's pointing placement, $F(1, 23) = 0.64$, $p = 0.43$, partial $\eta^2 = 0.03$. These results suggest that the observed efficiency gains in task completion time cannot be explained solely by changes in the instructor's verbal output.

3.3 Gesture duration made by the instructor

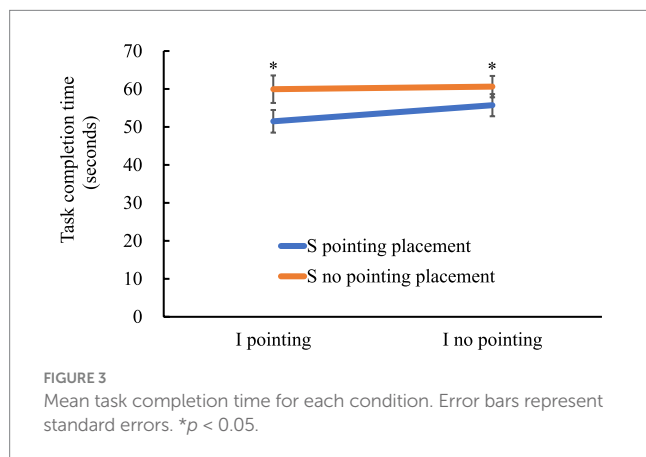
To investigate how the searcher's pointing behaviour influenced the instructor's gestural behaviour, we analysed the total gesture duration (in seconds) produced by the instructor per trial. Gesture durations were calculated using a segmented annotation approach, summing the time between the onset of gesture preparation and the end of retraction across all gestures in each trial (see Section 2.5).

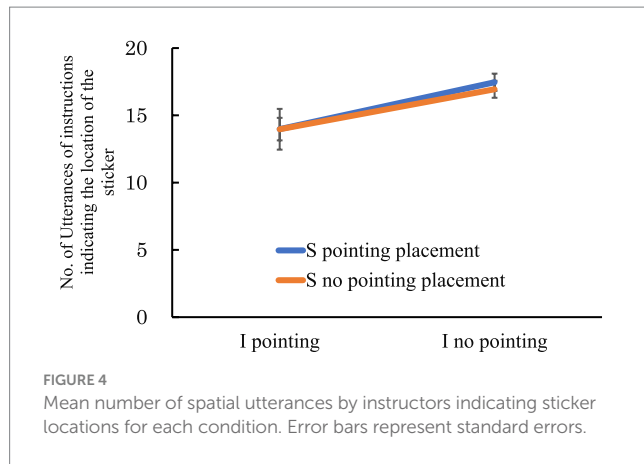
In this analysis, we focused on two conditions in which the instructor was allowed to use gestures. First, gesture duration was standardised by dividing it by the total trial duration. Then, for the standardised gesture durations, the mean for each condition was calculated per dyad. The overall mean across all dyads was then computed. In trials where searchers began with a pointing gesture (Searcher Pointing Placement—Instructor Pointing), instructors spent on average 52.7% of the trial time gesturing, which corresponded to 31.6 s of gesturing per minute of trial time (SD = 11.9). In contrast, in the Searcher No Pointing Placement—Instructor Pointing condition, the mean proportion of time gesturing was 44.2%, corresponding to 26.5 s per minute (SD = 12.0). A paired-samples t -test showed that instructors spent a significantly larger proportion of the trial gesturing in the pointing placement condition than in the no pointing placement condition, $t(23) = 2.67$, $p = 0.013$, Cohen's $d_z = 0.55$.

4 Discussion

4.1 Summary of findings

The present study examined how listeners' pointing gestures, defined as the initial placement of the searcher's index finger, affect the efficiency and dynamics of a collaborative spatial search task. Specifically, we aimed to determine whether the searchers' pointing gestures function not only as indicators of comprehension but also as visible resources that actively shape the instructor's communicative strategies. The findings produced two main outcomes. First, trials involving searchers' pointing placement resulted in significantly faster task completion times compared to those without such gestures. This suggests that listeners' pointing can contribute to the early alignment of shared spatial reference frames, reducing ambiguity in the instructor's guidance. Second, we observed that searchers' pointing placement elicited significantly longer gesture durations by instructors,





indicating that the presence of visible listener gestures may influence the production and sustainment of multimodal speaker behaviour. While our first hypothesis that instructors' pointing gestures would improve task efficiency was not supported, our second hypothesis was supported in that searchers' pointing facilitated more efficient collaboration and shaped the instructor's gestural behaviour. It is important to note, however, that the absence of a statistically significant effect of instructors' gestures should not be interpreted as definitive evidence of no effect. Rather, our analyses provide no evidence for such an effect under the current design, and it remains possible that smaller effects were not detected due to limited statistical power or the analytical constraints of dyad-level analyses.

Instructors' gestures did not enhance task efficiency or significantly reduce the number of utterances. This finding contrasts with previous research (e.g., [Bangerter and Chevalley, 2007](#)), which showed that pointing gestures facilitate communication in collaborative tasks by reducing verbal effort. One possible explanation lies in the communicative function of the instructor's pointing gestures. In [Bangerter and Chevalley's \(2007\)](#) study, instructors were free to use speech and/or gestures to directly identify each target, so gestures could replace parts of the verbal description. In contrast, in our study instructors were not allowed to point directly to the target dot on the searcher's board. As a result, gestures in our task mainly functioned as supportive cues rather than direct identifiers, leading to a smaller impact on verbal effort. This suggests that even when gestures are of the same type, such as pointing, their communicative function within the task context determines how strongly they interact with speech to affect efficiency.

4.2 Interpretation of listener gesture effects

The observed facilitative effect of searchers' pointing gestures on task efficiency can be interpreted in light of prior research emphasising the facilitative role of listeners' gestures in collaborative interaction. One possibility is that searchers' pointing acts as a visible cue of attentional and cognitive focus, enabling instructors to adapt their verbal and gestural strategies in real time. This interpretation aligns with the findings of [Healey et al.'s \(2015\)](#) study, which demonstrated that listener gestures actively contribute to maintaining mutual understanding, rather than functioning merely as feedback to the speaker. In our study, even the act of holding a finger at the centre of the board may have signalled a preliminary spatial commitment or

point of reference, thereby allowing instructors to anchor their instructions more efficiently.

Furthermore, listeners' gestures may function as a form of incremental feedback, as suggested by [Holler and Wilkin \(2011\)](#). They argued that gestural mimicry by listeners serves to display ongoing understanding and to facilitate the formation of common ground. In our context, while the searchers' pointing did not directly mimic the instructor's gestures, their presence nevertheless provided instructors with an embodied frame of reference, one that could be manipulated, verified or corrected. Such interpretations are supported by [Hosoma et al. \(2004\)](#), who reported that instructors in a similar spatial task spontaneously used the position of the listener's pointing finger to guide subsequent instructions. Our findings provide the first quantitative evidence for this phenomenon, demonstrating that searchers' pointing gestures are not merely expressions of understanding, but also serve as communicative resources that instructors actively use to organise and accelerate coordination.

4.3 Influence on speaker behaviour

In addition to its effect on task efficiency, searchers' pointing gestures also had a measurable influence on the gestural behaviour of instructors. Specifically, we observed that gesture durations were significantly longer in trials where searchers initiated the task with pointing. This finding suggests that the presence of searchers' visible bodily engagement not only facilitates listener understanding but also prompts instructors to modify and potentially extend their gestural expressions. This observation resonates with [Streeck's \(2017\)](#) ethnographic work, which described how co-participants reuse or elaborate on each other's gestures as a means of achieving alignment. In our study, longer gesture durations may reflect instructors' efforts to adapt to the spatial cues provided by the searcher, refining their own gestures to respond to the listener's evolving hand position and exploratory actions. Moreover, the dynamic quality of gesture in response to listener actions supports the notion that gestures are not static referential tools, but mutually elaborated actions situated within the material and interactional context, as articulated by [Goodwin \(2007\)](#). The extended duration of gestures may signal a shift from simple pointing to more sustained instructional behaviour.

4.4 Implications

The findings in the current study contribute to a growing body of research that reconceptualises gesture as an interactive, distributed process rather than a unilateral act of expression. By demonstrating that listeners' pointing gestures can significantly influence both the efficiency and structure of speaker behaviour, this study reinforces the view that gestures serve not only as referential acts but also as interactional scaffolds that shape how joint activities unfold. This interpretation aligns with theoretical perspectives that emphasise the reciprocal nature of gesture production and perception in real-time dialogue, where bodily actions are both responsive to and constitutive of shared understanding (e.g., [Clark, 1996](#); [Goodwin, 2007](#); [Streeck, 2017](#); [Schubotz et al., 2019](#)).

From an applied standpoint, the results also hold relevance for the design of collaborative systems and educational interfaces, where

real-time gestural input, particularly from non-speaking participants, can be used to inform instructional strategies or adaptive system responses. For example, in remote or virtual teamwork scenarios, the ability to interpret listeners' pointing gestures as indicators of comprehension or focus could facilitate more effective coordination. The present findings thus suggest that listeners' gestures should be treated as informationally rich signals, capable of shaping not only the flow of conversation but also the underlying structure of participation and guidance in joint tasks.

In summary, this study provides empirical support for the idea that listeners' pointing gestures, specifically searchers' initial finger placements, can function as active and interpretable resources in collaborative tasks. By showing that such gestures facilitate faster task completion and elicit more sustained gestural responses from instructors, we highlight the mutual, embodied coordination that underpins successful joint action. These results move beyond abstract notions of "common ground" by offering a more behaviourally grounded account of interactional alignment, centred on observable bodily actions and their real-time consequences.

4.5 Limitations

While the study provides valuable insights, several limitations should be acknowledged. The experimental setup involved a relatively constrained and task-specific environment, which may limit the generalisability of the findings to more naturalistic conversational contexts. Additionally, while gesture duration and utterance count provided quantitative indices of coordination, further qualitative analysis, such as fine-grained gesture type coding or sequential multimodal interaction analysis, could offer deeper insight into how gestures function moment-to-moment.

Most importantly, the experimental manipulation of searchers' pointing gestures was not entirely symmetrical with that of the instructors' pointing gestures. In the "with pointing placement" condition, the searcher began with her finger raised and kept it up, whereas in the "no pointing placement" condition, she started with her hands down and lowered her finger after each sticker was found. Although instructors were instructed to continue giving directions during this lowering phase, they often waited instead, potentially introducing variable "waiting times" that may have influenced task duration. Conversely, in the pointing placement condition, the searcher's sustained finger position sometimes obstructed the instructor's view of the board, leading to brief delays in identifying the next sticker. These unintended procedural effects introduce interpretive complexity, suggesting that the current manipulation may not fully isolate the effect of listeners' gestures *per se*.

Another source of variation relates to the initial strategies adopted by instructors when providing the first instruction for each sticker. In the pointing placement condition, instructors frequently issued gesture-manipulating instructions, such as "go that way from there," using the listener's finger as a reference point. In contrast, in the no-pointing placement condition, instructors often gave either precise spatial coordinates or general prompts like "place your finger somewhere," effectively prompting the listener to gesture. These differences in instructional format may have shaped the timing and frequency of instructors' gestures. Moreover, in a time-constrained setting, instructors might have

deliberately chosen to manipulate the listener's finger placement rather than describe sticker locations directly, thereby optimizing for efficiency. However, these communicative strategies were not controlled or measured in the current design. Future research should explore the decision-making processes of instructors through supplementary methods such as protocol analysis or retrospective interviews.

A further limitation is that our measure of instructional efficiency focused primarily on speech-based utterances. However, in many cases instructors also relied on non-verbal means, such as sustained or repeated pointing gestures without accompanying speech, to guide the searcher. These silent instructions could have reduced the number of verbal clauses required, thereby masking potential differences in utterance counts across conditions. Future studies should therefore incorporate both verbal and non-verbal measures of instruction to more comprehensively capture the multimodal nature of collaborative communication.

An additional limitation concerns the analytical approach adopted. A limitation of the present study is that our analyses relied on dyad-level repeated-measures ANOVAs. Although this approach is consistent with previous work on gesture and collaborative tasks (Bangerter and Chevalley, 2007; Kraut et al., 2003; Wang et al., 2021), it assumes that the dyad is a stable analytical unit across all conditions. In reality, the instructor and searcher roles alternated between individuals, creating a nested structure in which participants are embedded within dyads. While we chose ANOVA to avoid the added complexity of modelling these nested random effects, we note that Linear Mixed Models (LMMs) provide a more flexible alternative. Indeed, our exploratory LMM analyses produced the same pattern of results as the ANOVA, suggesting that our conclusions are robust. Future studies should nonetheless consider applying LMMs to more fully capture the hierarchical and role-switching nature of such collaborative tasks.

Finally, while our analyses focused on the instructors' speech and gestures, we did not examine the searchers' speech patterns in detail. It is possible that differences in trial duration partly reflected extended speech phases or clarification questions from the searchers. Future studies should therefore include systematic analyses of listener speech to provide a more comprehensive account of multimodal coordination.

Despite these limitations, the current study makes a significant contribution to gesture research by emphasising the role of listeners' bodily actions not as passive reflections of understanding, but as active, interpretable components of collaborative discourse.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Waseda University (Approval No. 2023–125). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

KS: Conceptualization, Writing – review & editing, Supervision. KoK: Data curation, Investigation, Writing – review & editing, Methodology, Writing – original draft. KeK: Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Bangerter, A., and Chevalley, E. (2007). Pointing and describing in referential communication: when are pointing gestures used to communicate? SluisVan Der, E. Reiter and E. Krahmer Proceedings of the workshop on multimodal output generation (MOG 2007) (pp. 17–28).
- Bentley, B., Walters, K., and Yates, G. C. R. (2023). Using iconic hand gestures in teaching a year 8 science lesson. *Appl. Cogn. Psychol.* 37, 496–506. doi: 10.1002/acp.4052
- Clark, H. H. (1996). Using language. Cambridge, UK: Cambridge University Press.
- Goodwin, C. (2007). Participation, stance and affect in the organization of activities. *Discourse Soc.* 18, 53–73. doi: 10.1177/0957926507069457
- Healey, P. G., Plant, N. J., Howes, C., and Lavelle, M. (2015). When words fail: Collaborative gestures during clarification dialogues. In turn-taking and coordination in human-machine interaction: Papers from the AAAI spring symposium (pp. 23–29)
- Holler, J., Kendrick, K. H., and Levinson, S. C. (2018). Processing language in face-to-face conversation: questions with gestures get faster responses. *Psychon. Bull. Rev.* 25, 1900–1908. doi: 10.3758/s13423-017-1363-z
- Holler, J., and Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *J. Nonverb. Behav.* 35, 133–153. doi: 10.1007/s10919-011-0105-6
- Hosoma, H., Ishizu, K., Shigematsu, M., Nakamura, T., and Yano, M. (2004). Conversations that show each other's body: Talking about the others body with self-body. Proceedings of the 14th annual meeting of the Association of Sociolinguistic Sciences, 62–81
- Kang, S. H., and Tversky, B. (2016). From hands to minds: gesture promotes understanding. *Cogn. Res. Princ. Implic.* 1, 1–9. doi: 10.1186/s41235-016-0004-9
- Kendon, A. (2004). Gesture: Visible action as utterance: Cambridge University Press.
- Kimbara, I. (2006). On gestural mimicry. *Gesture* 6, 39–61. doi: 10.1075/gest.6.1.04kim
- Kraut, R. E., Gergle, D., and Fussell, S. R. (2003). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. CSCW '02: Proceedings of the 2002 ACM conference on computer supported cooperative work 31–40
- Lausberg, H., and Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behav. Res. Methods* 41, 841–849. doi: 10.3758/BRM.41.3.841
- McNeill, D. (1992). Hand and mind: What gestures reveal about thought. Chicago, IL: University of Chicago Press.
- Schubotz, L., Özyürek, A., and Holler, J. (2019). Age-related differences in multimodal recipient design. *Lang. Cogn. Neurosci.* 34, 254–271. doi: 10.1080/23273798.2018.1527377
- Sekine, K., and Özyürek, A. (2024). Children benefit from gestures to understand degraded speech but to a lesser extent than adults. *Front. Psychol.* 14:1305562. doi: 10.3389/fpsyg.2023.1305562
- Streeck, J. (2017). Self-making man: A day of action, life, and language. Cambridge, UK: Cambridge University Press.
- Wang, I., Narayana, P., Patil, D., Bangar, R., Draper, B. A., Beveridge, J. R., et al. (2021). "It's a joint effort: understanding speech and gesture in collaborative tasks" in Human-Computer Interaction Techniques and Novel Applications (HCII 2021). ed. M. Kurosu, 159.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.