AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 9, 2623–2635
https://doi.org/10.1037/xge0001402

Gestures Cued by Demonstratives in Speech Guide Listeners' Visual Attention During Spatial Language Comprehension

Demet Özer[1, 2], Dilay Z. Karadöller[1, 3], Aslı Özyürek[3, 4], and Tilbe Göksun[1]

[1] Department of Psychology, Koç University

[2] Department of Psychology, Kadir Has University

[3] Multimodal Language Department of Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

[4] Donders Institute for Brain, Cognition and Behavior, Radboud University

Gestures help speakers and listeners during communication and thinking, particularly for visual-spatial information. Speakers tend to use gestures to complement the accompanying spoken deictic constructions, such as demonstratives, when communicating spatial information (e.g., saying "The candle is *here*" and gesturing to the right side to express that the candle is on the speaker's right). Visual information conveyed by gestures enhances listeners' comprehension. Whether and how listeners allocate overt visual attention to gestures in different speech contexts is mostly unknown. We asked if (a) listeners gazed at gestures more when they complement demonstratives in speech ("*here*") compared to when they express redundant information to speech (e.g., "*right*") and (b) gazing at gestures related to listeners' information uptake from those gestures. We demonstrated that listeners fixated gestures more when they expressed complementary than redundant information in the accompanying speech. Moreover, overt visual attention to gestures did not predict listeners' comprehension. These results suggest that the heightened communicative value of gestures as signaled by external cues, such as demonstratives, guides listeners' visual attention to gestures. However, overt visual attention does not seem to be necessary to extract the cued information from the multimodal message.

Public Significance Statement

Natural face-to-face human communication involves many multimodal cues, such as co-speech hand gestures. This study investigates listeners' visual attention to gestures that are communicatively cued by speech (i.e., demonstratives) during the comprehension of visual-spatial language. This study suggests that how speakers design their multimodal utterances guides listeners' language processing.

Keywords: overt visual attention to gestures, gesture processing, spatial language, gestures with demonstratives

Supplemental materials: https://doi.org/10.1037/xge0001402.supp

This article was published Online First April 24, 2023.

Demet Özer https://orcid.org/0000-0003-3230-2874

We thank İrem Türkmen for her assistance with data collection, Gamze Turunç, and Zeynep Aslan for their help in stimuli preparation. We are grateful for the helpful feedback we received from the Language & Cognition Lab members at Koç University, Istanbul, and Multimodal Language and Cognition Lab at Max Planck Institute for Psycholinguistics, Nijmegen.

This work was supported by the TÜBİTAK's (The Scientific and Technological Research Council of Turkey) International Research Fellowship Programme for PhD Students (2214-A) given to Demet Özer, Türkiye Bilimler Akademisi (Turkish Academy of Sciences) Outstanding Young Scientist Award 2018 and a James McDonnell Foundation Scholar Award (Grant 220020510) given to Tilbe Göksun.

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

Part of the data and ideas were presented as an online poster presentation at the Seventh Gesture and Speech in Interaction (GESPIN) conference on September 2020.

The data presented in the current manuscript can be found online in the Open Science Framework repository at https://osf.io/vkhf5/?view_only=cf38e23f09a74737999385a1b46de0e2.

Demet Özer served as lead for data curation, formal analysis, visualization, and writing–original draft. Dilay Z. Karadöller served in a supporting role for conceptualization, methodology, project administration, and writing–review and editing. Aslı Özyürek served as lead for supervision and contributed equally to resources, and Tilbe Göksun served as lead for funding acquisition, resources, and supervision. Demet Özer, Aslı Özyürek, and Tilbe Göksun contributed to conceptualization, project administration, writing–review and editing, and methodology equally.

Correspondence concerning this article should be addressed to Demet Özer, Department of Psychology, Kadir Has University, Kadir Has Caddesi Fatih, 34083 Istanbul, Turkey. Email: demet.ozer@khas.edu.tr

Individuals use their hands to think and communicate (McNeill, 1992), particularly for spatial information, such as relative spatial relations between objects (e.g., *left–right* or *on–under*; Alibali, 2005; Beattie & Shovelton, 1999a, 2002a; Dargue et al., 2019; Hostetter, 2011; Karadöller et al., 2019; Karadöller, Sümer, Ünal, & Özyürek, 2021; Lavergne & Kimura, 1987). Speakers tend to use gestures to complement spoken deictic constructions, such as place-referring demonstratives (e.g., "*here*") when describing spatial relations (e.g., Beattie & Shovelton, 2006; Holler & Stevens, 2007). For example, a speaker may say, "The candle is *here*," and gesture to the right side to express that the candle is on the speaker's right (e.g., Cooperrider, 2017; Emmorey & Casey, 2001). In turn, listeners attend to, process, and benefit from observing those gestures when comprehending objects' relative position (Beattie & Shovelton, 1999a, 1999b, 2002a, 2002b; Holler et al., 2009). However, it is less known about contextual modulation of the relation between speech and gesture on comprehension. The current study focuses on how listeners process gestures when they are used with demonstratives in a complementary fashion and the demonstrative in speech cues the information in gesture versus not. To answer this, we measured listeners' direct visual attention to gestures by employing an eye-tracking paradigm. We investigated listeners' comprehension and allocation of direct attention to gestures during the comprehension of spatial relations between objects in different speech contexts. In particular, we examined visual attention to gestures that expressed redundant versus complementary information to the speech through the use of demonstratives, which potentially cue gestures to be central for successful communication.

Listeners' Visual Attention to Gestures

In face-to-face interaction, addressees mainly look at their interlocutors' faces, given that the maintained mutual gaze between interlocutors is a social norm (Argyle & Cook, 1976; Argyle & Graham, 1976; Gobel et al., 2015). Eye-tracking research showed that listeners/viewers spent approximately 90% of the total viewing time fixating speakers' faces (i.e., primarily the nose bridge and the eye area) and fixated only a minority of speakers' gestures in both live face-to-face and video interactions (Beattie et al., 2010; Gullberg & Holmqvist, 1999, 2006; Gullberg & Kita, 2009; but see Nobe et al., 1997, 2000 for increased overt gaze allocation to gestures when interacting with an anthropomorphic agent).

However, several kinematic features of gestures modulate listeners' overt gaze allocation. For example, Gullberg and Holmqvist (1999) showed that listeners fixated gestures more when they were executed in the peripheral gesture space (e.g., away from the body/torso) compared to ones that were performed in the central gesture space (e.g., closer to face and body/torso), particularly for the vertical axis (but see Gullberg & Kita, 2009 for no effect of gesture space). Additionally, gestures with slower strokes (i.e., the most meaningful part of the gesture) and longer poststroke holds attracted more fixations (Gullberg & Holmqvist, 2006; Gullberg & Kita, 2009; Nobe et al., 1997, 2000).

Apart from kinematic features, another factor that might affect listeners' visual attention to gestures is the relative communicative/informative value of gestures in relation to speech. That is, listeners might actively seek alternative sources of disambiguating information in the case of disfluent, noisy, or nonnative speech comprehension. Thus, when speech is under informative, listeners might attend more to other communicative cues, such as gestures. Indeed, studies suggest that the degree of language experience modulates how much listeners attend to gestures (e.g., Drijvers et al., 2019; Drijvers & Özyürek, 2017, 2018; Rimé et al., 1988; but see Gullberg & Holmqvist, 1999). Listeners fixated gestures more when watching narratives told in a partially comprehensible (i.e., French) or completely incomprehensible language (i.e., Russian) compared to when watching narratives told in their native language (i.e., Flemish; Rimé et al., 1988). In a recent study, Drijvers et al. (2019) also showed that nonnative listeners looked at gestures more than native listeners when comprehending Dutch action verbs both in noisy and clear speech. Moreover, they found that native listeners benefited from fixating gestures when comprehending noisy speech, suggesting that listeners might attend gestures as an alternative source of information to resolve insufficiencies in noisy speech.

Related to this, Yeo and Alibali (2017) investigated listeners' visual attention to gestures that expressed information necessary to disambiguate fluent versus disfluent speech (i.e., speech with filled pauses "*um*"). In Yeo and Alibali (2017), the actress said, "The triangle changed color," and made a gesture of an upward-pointing triangle followed by a shape array containing either both upward- and downward-pointing triangles (i.e., nonredundant gesture) or only an upward-pointing triangle (i.e., redundant gesture [RG]). The total fixation duration to gestures did not differ across redundant versus nonredundant gestures. However, when analyzed whether listeners fixated gestures at least once in a dichotomous way (0–1), they found that listeners tended to fixate gestures more when they expressed nonredundant than redundant information. Moreover, they tended to gaze at gestures more when speech was disfluent than fluent (Yeo & Alibali, 2017). Also, in a recent eye-tracking study, van Nispen et al. (2022) asked neurotypical listeners to watch video clips in which speakers with and without aphasia described a vivid scenario. They found that listeners gazed more at gestures that were produced by speakers with aphasia who had less informative speech than to gestures that were produced by speakers without aphasia. Overall, these results suggest that gestures might have heightened communicative value for listeners when there are certain insufficiencies in speech, which guides listeners' visual attention to those gestures.

Gesture's communicative/informational value could also be modulated by the properties of speakers' utterance design as gestures might be cued to be informative by the speaker. As outlined in Cooperrider (2017), speakers might "foreground" their gestures when gestures are helpful to convey some critical aspect of the message and design their multimodal packages accordingly. In such cases, speakers signal to the listener that gestures are central for successful communication with different accompanying cues. Listeners, in turn, attend to gestures more as they are cued to be central by the speaker. The heightened communicative value of gestures in such cases might guide listeners' visual attention to gestures. For example, Gullberg and colleagues (Gullberg & Holmqvist, 1999; Gullberg & Kita, 2009; but see Gullberg & Holmqvist, 2006) examined listeners' visual attention to gestures when gestures are cued to be central by speakers' fixations on their own gestures (i.e., auto-fixations). They found that listeners gazed at gestures more during auto-fixations, suggesting that visual cues (e.g., speaker's gaze) could mark the relevance of the gesture by establishing shared attention (Enfield, 2009) and guide listeners' attention to gestures.

Another hallmark of such foregrounding gestures is the concurrent use of demonstratives in speech (Cooperrider, 2017). Listeners might attend to gestures as they are cued to be central to the message when they are co-produced with demonstratives. To our knowledge, there is no empirical research on listeners' attention to gestures when they are used along with demonstratives.

A follow-up question to whether and how gestures that are cued to be communicative (i.e., with demonstratives) attract gaze fixations if these fixations enhance information uptake from gestures. Previous studies suggested that there is no relation between gazing at gestures and semantic uptake from gestures as gestures can be processed peripherally (e.g., Beattie et al., 2010; Gullberg & Kita, 2009). For example, Gullberg and Kita (2009) found that listeners were more likely to uptake information from gestures that were initially fixated by speakers themselves compared to gestures that were not. However, when listeners' own fixations to those gestures were analyzed (auto-fixations), there was no relation between gazing at gestures and information uptake. These findings are also in line with the sign language literature, showing that peripheral perception is sufficient for signs to be processed in parallel with overt visual attention to the face (e.g., Emmorey et al., 2000). However, Drijvers et al. (2019) showed that only native, but not nonnative listeners' gazing at gestures predicted how much they benefited from observing gestures during degraded speech comprehension. These findings suggest that although nonnative listeners gazed at gestures more than native listeners, they might be more hindered by degraded speech and need more cues (i.e., phonological cues in visible speech) to aid comprehension during degraded speech (Drijvers & Özyürek, 2017, 2018). This suggests that the relation between listeners' fixations on gestures and comprehension of the message is modulated by the factors in speech accessibility. We do not know, however, how other factors, such as the relation between gesture and explicit cues in speech modulate this relation.

To answer these remaining questions regarding the influence of communicatively cued gestures by speech on visual attention and information uptake, we ask whether and how (a) listeners fixate gestures that are co-produced with place-referring demonstratives ("*here*") compared to ones with redundant speech (e.g., "*right*") and (b) gazing at gestures modulates comprehension of the multimodal message with different speech and gesture relations.

Gestures in Spatial Language

Gestures have a unique role in spatial language and cognition (Alibali, 2005; Dargue et al., 2019; Hostetter, 2011). Gestures help both speakers and listeners when expressing, communicating, and thinking about spatial information (e.g., Allen, 2003; Chu & Kita, 2008, 2011; Emmorey et al., 2000; Göksun, Goldin-Meadow, et al., 2013; Hostetter et al., 2011; Özer et al., 2017; So et al., 2015). One spatial context in which speakers use an abundant number of gestures is the descriptions of relative spatial relations (e.g., left–right, on–under, front–behind, next to) between a figure object (i.e., the object whose relative position to be located) and a ground object (i.e., the reference object; Driskell & Radtke, 2003; Göksun, Lehet, et al., 2013; Holler et al., 2009; Karadöller et al., 2019; Karadöller, Sümer, Ünal, & Özyürek, 2021; Karadöller, Sümer, & Özyürek, 2021; McNeil et al., 2000).

Given gesture's expressive potential, speakers often convey visual-spatial information such as object size, object location, manner of movement, and spatial location in gestures rather than in speech (Beattie & Shovelton, 2006; Emmorey & Casey, 2001; Holler & Stevens, 2007; Karadöller et al., 2019, Karadöller, Sümer, & Özyürek, 2021; Kita & Özyürek, 2003). Speakers tend to encode spatial information such as object's size only in their gestures rather than in speech, especially when the size information is crucial (Beattie & Shovelton, 2006) or when they talk to interlocutors with whom the size information is new (Holler & Stevens, 2007). This suggests that the complementary distribution of information expressed by gesture versus speech lies on a continuum during utterance production (Goldin-Meadow, 2003). Gestures, especially for spatial language, often complement and provide additional meaning to spoken expressions on a semantic level (e.g., Beattie & Shovelton, 2006; de Ruiter et al., 2012; Gerwing & Allison, 2009; Melinger & Levelt, 2005).

Complementing this, studies also showed that observing gestures enhance listeners' comprehension of spatial information, mainly when they express some semantic features such as size, shape, and the relative position of objects (Beattie & Shovelton, 1999a, 1999b, 2002a; Holler et al., 2009; Hostetter, 2011). The current study is on the processing of gestures during the comprehension of spatial relations between objects.

One instance in which gestures are often employed to complement accompanying speech is the use of spoken deictic constructions (i.e., demonstratives) in spatial language. Speakers tend to use gestures along with demonstratives (e.g., "like this" or "here") that marks the relevance and the importance of their gestures for the communication (e.g., Emmorey & Casey, 2001; Slonimska et al., 2015). The accompanying demonstratives in speech refer to speakers' gestures and signals to listeners that an essential part of the message is conveyed through the gesture channel. For example, a speaker who says, "The candle is *here*," can produce a gesture that shows the right side of the speaker to express that the candle is on the speaker's right side (Karadöller et al., 2019, 2022). In such a multimodal utterance, the gesture expresses the critical complementary information to the accompanying speech and is communicatively cued by speech for the successful comprehension. We do not know however whether such contexts enhance attention to such gestures and information uptake from them.

The Present Study

Following on from these studies, we examined whether and how listeners allocated direct visual attention to gestures that are co-produced with different speech contexts. More specifically, we asked (a) whether listeners gazed at gestures more when they expressed critical complementary information (i.e., when gestures are coupled and thus cued communicatively with demonstratives) compared to redundant information in relation to the accompanying speech, and (b) whether and how gazing at gestures modulated how much listeners benefited from observing gestures during the comprehension of relative spatial relations. To this end, we used eye-tracking to record participants' eye movements, which reflect attention allocation to gestural information (Posner, 2016). We also tested listeners' comprehension of the multimodal message with a forced-choice task for pictures that depicted spatial relations between objects.

Although research suggests that observing gestures facilitates listeners' comprehension of spatial relations (e.g., Beattie & Shovelton, 1999a, 1999b, 2002a; Holler et al., 2009), the facilitative effects of observing gestures might be different for different types

of spatial relations. Therefore, we tested the comprehension of two types of spatial relations: viewpoint-dependent (*left–right*) and viewpoint-independent (*on–under*). *Viewpoint-dependent* spatial relations such as *left–right* require viewpoint alignment and pose a challenge to both speakers and listeners compared to *viewpoint-independent* spatial relations such as *on–under* (Galati & Avraamides, 2013; Galati et al., 2013; Karadöller, 2022; Keysar et al., 2000). The gestural expression of left–right directionality poses a viewpoint coordination problem as left–right gestures that were produced by the speaker's egocentric perspective creates an incongruence between the veridical position in which the gesture is made and the intended meaning of it (e.g., Miller & Johnson-Laird, 1976). Indeed, gestures that expressed left–right directionality hindered listeners' comprehension (Gullberg & Kita, 2009; Hostetter et al., 2018; Pyers et al., 2015). Therefore, in the current study, we also tested listeners' comprehension and visual attention to gestures during spatial language across different spatial relations that require viewpoint alignment or not.

To investigate these questions, we employed an offline comprehension paradigm in which we presented participants with short video clips of an actress describing either viewpoint-dependent (*left–right*) or viewpoint-independent (*on–under*) spatial relations between two objects in three different conditions: (a) only in speech without making any gesture (speech-only [SO], e.g., saying *right* while standing still), (b) both in speech and in gesture (RG, e.g., saying *right* while gesturing to the right side), and (c) only in gesture with a demonstrative in speech (complementary gesture [CG], e.g., saying *here* while gesturing to the right side). The actress spoke two sentences: the first sentence introduced the figure and the ground objects (e.g., "There is a vase and a candle"), the second sentence introduced the relative spatial relation of the figure object (e.g., "The candle is on the *right*"). The actress made the gesture during the second sentence preceding the spatial term. After watching the videos while their eye gaze was measured, participants were asked to choose the picture that best depicts the described spatial relation among four alternatives to measure their comprehension. We were interested in accuracy, reaction times (RTs), and the percent number of fixations to the gestural space, particularly during the second sentence, which included target information.

For comprehension, we predicted that:

(1) Performance would be the worst (the lowest accuracy and the longest RTs) when the critical spatial information was expressed only in gesture (i.e., CG) compared to RG and SO conditions.
(2) There would be a difference between RG and SO conditions in regard to the type of the spatial relation. For *on–under*, participants would be more accurate and faster in RG than SO condition (Beattie & Shovelton, 1999a; Holler et al., 2009). For *left–right*, participants would be less accurate and slower in RG than in SO condition (Hostetter et al., 2018; Pyers et al., 2015).

For eye gaze, we predicted that:

(3) Listeners would fixate gestures more in the CG condition than in the RG condition for both types of spatial relations. Moreover, this effect could be stronger for left–right than for on–under.

For the interaction between eye gaze and comprehension, we predicted that:

(4) Gazing at gestures would be related to enhanced comprehension. Listeners would be more accurate and faster as they fixate gestures more. Yet, this effect would be more salient for CG condition than for RG condition, in which the critical spatial information can be extracted from both channels. Also, this effect could be stronger for left–right than for on–under.

Method

Participants

We recruited 51 native Turkish speakers (36 females, $M_{age} = 21.2$ years, $SD_{age} = 2.6$, age range = 18–30 years, $M_{education} = 15.4$ years, $SD_{education} = 2.55$) from Koç University, İstanbul in return for either course credit or monetary compensation. All participants were right-handed, had normal or corrected-to-normal vision, normal hearing, and were not taking any psychiatric or neurological medication at least 6 months before the testing. All participants gave informed consent before the experiment, which was approved by the Institutional Review Panel for Human Subjects of Koç University.

Materials

The experimental task consisted of short video clips of an actress describing either viewpoint-dependent (i.e., *left–right*) or viewpoint-independent (i.e., *on–under*) spatial relations between two objects. In these clips, the actress uttered two sentences and sometimes made gestures to describe the location of the figure object in three different conditions: (a) In the *speech-only condition* (*SO*), the actress uttered the sentence with the specific spatial term without any hand gesture (e.g., the actress said that "There is a vase and a candle. The candle is on the *right*" while standing still), (b) in the *redundant-gesture condition* (*RG*), the actress uttered the sentence with the specific spatial term and made a gesture that showed the location of the figure object (e.g., the actress said that "There is a vase and a candle. The candle is on the *right*" while showing her right side), and (c) in the *complementary-gesture condition* (*CG*), the actress uttered the sentence with a demonstrative (i.e., *here*) and made a gesture to show the location of the figure object (e.g., the actress said that "There is a vase and a candle. The candle is *here*" while showing her right side to express that the candle is on the right of the vase). Figure 1 depicts the different conditions in the experiment.

All videos displayed the actress from the head to the knees, appearing in the same starting position (i.e., in the middle of the screen) with hands casually hanging on each side of the body. The actress wore black clothes, and the background was white. In each clip, the actress uttered two sentences: the first sentence introduced the ground and the figure objects, respectively (e.g., "There is a vase and a candle"), whereas the second sentence introduced the relative spatial location of the figure object (e.g., "The candle is on the *left/right/on/under/in*" or "The candle is *here*"). For our speech stimuli, we made use of an earlier study in which Turkish-speaking adults were asked to spontaneously describe *left–right* and *on–under* spatial relations between a central ground object and figure object

Figure 1

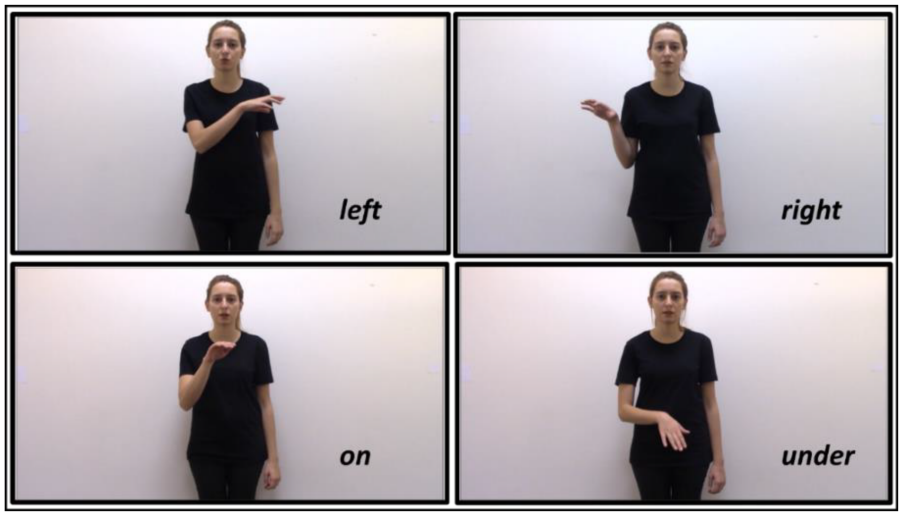
Three Conditions in the Experimental Task

Note. Although written in English on the images, the stimuli originally used in the study were in Turkish. The underlined word denotes the speech that gesture temporally overlaps with.
*Vazo ve mum var. Mum *sağ-da /**bur-da.*
Vase and candle there_is. Candle *right-LOC / here-LOC.*
GROUND FIGURE FIGURE SPATIAL RELATION
'There is a vase and a candle. The candle is on the right.'
See the online article for the color version of this figure.

(the same set of pictures-to-be-described in Karadöller et al., 2019 study was used as pictures for the response screen in the current study, see also Karadöller, 2022). In Karadöller et al. (2019) study, we examined Turkish-speaking adults' speech on the use of the figure object, the ground object, and the spatial relation when describing picture arrays. Based on this examination, we found that Turkish speakers mainly first introduced the ground object, followed by the figure object (e.g., "there is a vase [GR] and a candle [FIG]"). Then, they tend to present the figure object and its relative location to the ground object without necessarily naming the ground object (e.g., "the candle [FIG] is on the right"). To make our stimuli as natural as possible, we prepared the speech in the current study according to these findings of Karadöller et al. (2019). During the second sentence, the actress made the gestures with her right hand and retracted her arm back to the initial position. For *left–right* gestures, the actress extended her arm up to the side of her left/right arm. For on–under gestures, the actress made the gestures around her torso, right hand showing slightly above to the face for showing "*on*," and slightly below to the feet for "*under*." Figure 2 presents the left–right and the on–under gestures used in the experimental paradigm. For filler trials, we also included clips with "*in*." Participants saw videos in which the actress described the "*in*" spatial relation for all the three conditions (10 trials per condition, 30 filler trials in total). However, these trials were fillers and not used for

Figure 2

Left–Right and On–Under Gestures That Were Used in the Experimental Task



Note. See the online article for the color version of this figure.

the analyses. All videos were 5-s long. The critical target word (e.g., "*left*" or "*here*") started around 3,200 ms after the onset of the video. The onsets of gesture preparation and stroke were around 2,700 ms and 3,400 ms after the onset of the video. We inserted the audio files of SO videos (i.e., the videos in which the actress does not make any gesture) into the videos in which the actress used a gesture. When a speaker makes a gesture, the prosodic prominence of the gesture's referent is altered in speech (Krahmer & Swerts, 2007). We recorded SO versions for each trial with a gesture component. We then swapped the audio files to eliminate the possible confounding effects of this altered prosodic prominence among conditions.

Procedure

Participants were tested in a dimly lit soundproof room on a 17-in. Acer laptop on which a mouse, headphones, and eye-tracking device were connected. Participants were instructed to watch the short video clips and then choose the picture that best depicted the spatial relation described in the video among four alternatives. In the response screen, they were presented with four pictures showing different spatial relations between the same figure and the ground object. Ground objects were always in the center of the pictures, and the figure object's location in relation to the ground object changed in each picture. They were asked to choose the correct picture by clicking with the mouse as accurately and fast as possible (see Figure 3 for a sample response screen). In a canonical trial, participants were presented with a "Get Ready!" screen for 800 ms, followed by a brief preview of the response screen (i.e., picture array) for 500 ms to familiarize participants with the appearances of the figure and the ground objects before they listened to the spatial descriptions. Then, participants saw the fixation cross for 1,000 ms, followed by the video for 5,000 ms. After the clip, the self-timed response screen appeared, and the participants clicked on one of the pictures in the array with the mouse. Figure 3 depicts a sample trial. In the experimental task, we measured participants' accuracy (correct–incorrect) and RTs to respond, which were time-locked to the onset of the response screen as gesture strokes had slightly varying onsets across preceding videos.

Before starting the experimental task, participants went through familiarization and practice sessions. In the familiarization session, participants saw seven pictures depicting different spatial relations between two different objects and written labels for each spatial relation (i.e., *left*, *right*, *front*, *behind*, *on*, *under*, *in*; please note that front–behind were not used in the video stimuli, however, these spatial relations were depicted in the pictures as foils during the response screen). Later, the experimenter demonstrated the task with five examples for each spatial relation used in the experiment (i.e., *left–right, on–under, in*). After familiarization, the participants completed 12 practice trials and were given oral feedback (correct–incorrect) by the experimenter. After these trials, participants have seated approximately 70 cm away from the computer and were instructed to limit their movements during the session. We used Tobii Pro X3-120 to monitor eye movements at a sampling frequency of 120 Hz. Participants underwent a nine-point calibration and validation procedure. The calibration proceeded until the discrepancy between the calibration point and the participant's eye gaze was $<1°$.

The participants completed 150 trials in total, 10 for each spatial relation and condition: 10×5 spatial relations (*right*, *left*, *on*, *under*, *in*) $\times$ 3 conditions (*speech-only, redundant gesture, complementary gesture*). Out of 150 trials, 120 were experimental trials, and the rest ("*in*") were filler items. All trials were presented randomly in E-Prime 3.0 with Tobii extension. The experiment lasted approximately 30 min in total. The exemplar stimuli and the E-Prime code for the experimental task can be found in Open Science Framework repository at https://osf.io/vkhf5/?view_only=cf38e23f09a74737999385a1b46de0e2.

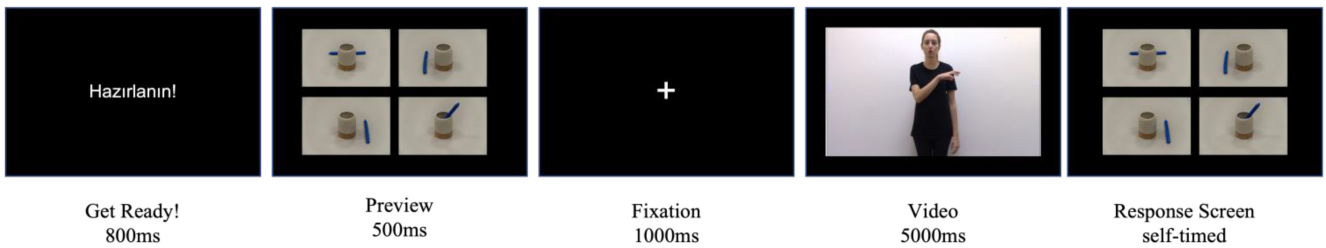
Analyses and Results

We only analyzed 120 experimental trials (excluding filler items of *in*) per participant. For analyses related to the task performance, there were 6,120 responses (120×51 participants). For RT, we discarded outliers that were two standard deviations above or below the mean ($n = 196$, 3.2%) and incorrect responses ($n = 438$, 7.16%). See Table S1 in the online supplemental materials for the number of analyzed items in each condition. For the eye-gaze data, we excluded two participants due to an experimental error for eye-tracking. We discarded trials with a track loss of 25% and higher and trials in which all the fixations were in non-AOI space for the time-of-interest ($n = 85$, 1.45%).

Eye-Tracking Analyses

The eye movements were automatically coded as fixations and saccades using Tobii Pro Lab algorithms with Tobii Pro I-VT Filter in which the velocity threshold is set to 100°/s. We defined two areas of interest (AOIs): face and gestural space (see Figure S1 in the online supplemental materials). The gestural space AOI started from the upper neck of the actress and ended at

Figure 3
A Sample Trial in the Experimental Task



Note. See the online article for the color version of this figure.

the hand wrists when her arms hung casually on her sides (see Drijvers et al., 2019 for a similar approach). We created the gestural space AOI length based on the *x*-coordinates that correspond to the furthest point in which a gesture was made on the screen. The size and the location of the AOIs were equal across videos. We measured the proportion of the number of fixations to the gestural space and the face AOIs only during the second sentence in the video, which included the target information and the gesture. Table S2 in the online supplemental materials shows the proportion of the number of fixations and the proportion of the total fixation duration to each AOI relative to the total number of fixations to the entire screen (i.e., AOI + non-AOI spaces) and the total time of the video, respectively. However, for analyses, we calculated the proportion of the number of fixations to the gestural space relative to the total number of fixations to the gestural space and the face AOIs (i.e., fixations to gestures divided by the sum of fixations to the gesture + face).

Mixed-Effects Analyses

We used multi-level mixed-effects models to incorporate all data variance in our analyses. We used the logistic version of the model for binary outcome variables (i.e., accuracy). In all models, the random effects included the random subject- and item-intercepts. We performed all analyses with *lme4* package (Bates et al., 2015) on R Studio (RStudio Team, 2020). In generalized (i.e., logistic) models, we used "bobyqa" optimizer that maximized the number of iterations performed in a model to alleviate possible convergence problems (Powell, 2009). We used *car* package (Fox & Weisberg, 2019) to obtain Type-III Wald Chi-Square Test results, which showed whether the inclusion of each term in a model (i.e., explanatory variables) significantly improved the model. We reported Bonferroni-adjusted pairwise comparisons by *emmeans* package in R (Lenth, 2021). We used *ggplot2* (Wickham, 2016) and *jtools* (Long, 2020) packages for data visualization. The R code and the datasets are available in the Open Science Framework repository at https://osf.io/vkhf5/?view_only=cf38e23f09a74737999385a1b46de0e2.

Results

Experimental Task Performance Results

First, we asked how task performance differed across conditions and spatial relation types. In Model 1, the outcome variable was accuracy, and in Model 2, the outcome variable was RT. The fixed factors included the main effects of the condition, the spatial relation type, and the two-way interaction between them. Descriptive statistics about accuracy and RT for each condition can be found in Table S1 in the online supplemental materials.

For accuracy (Model 1, see the left graph in Figure 4), there was a main effect of the condition, $\chi^2(2) = 69.74$, $p < .001$. CG condition was associated with lower accuracy by -1.11 ± 0.18 compared to SO condition ($z = -6.10$, $p < .001$) and by -1.42 ± 0.19 compared to RG condition, $z = -7.47$, $p < .001$. There was no difference between RG and SO conditions, $z = 1.49$, $p = .41$. There was also a main effect of the spatial relation type, $\chi^2(1) = 13.83$, $p < .001$. Across all conditions, on–under trials were associated with decreased accuracy by -0.52 ± 0.16 compared to left–right trials, $z = -3.34$, $p < .001$. There was no interaction between condition and spatial relation type, $\chi^2(2) = 2.13$, $p = .35$.

For RT (Model 2, see the right graph in Figure 4), there was no main effect of the condition, $\chi^2(2) = 5.61$, $p = .06$. Yet, there was a main effect of the spatial relation type on RTs, $\chi^2(1) = 31.82$, $p < .001$. For all conditions, on–under trials were associated with increased RTs by 156 ± 27.7 compared to left–right trials, $z = 5.63$, $p < .001$. There was no interaction between condition and spatial relation type, $\chi^2(2) = 0.36$, $p = .84$.

In sum, we found that participants have lower accuracy in CG condition compared to RG and SO conditions. Participants also have lower accuracy and slower RTs for on–under trials compared to left–right trials. There were no differences between RG and SO conditions for either accuracy or RT.

Eye-Tracking Results

Table S2 in the online supplemental materials presents descriptive statistics for the proportions of the total number of fixations and the fixation duration to the face and the gestural space AOIs across different conditions. Participants, on average, spent 23% of the total viewing time fixating to the gestural space and around 70% of the total viewing time fixating to the face.

We asked how the proportion of fixations to the gestural space during the second sentence[1] changed across conditions and spatial relation types (Model 3, see Figure 5). The fixed effects included the main effects of the condition, the spatial relation type, and the two-way interaction between them. Please note that although there were no gestures in the SO condition, we still measured eye gaze to the gestural space in this condition to have a baseline measure against the other two conditions with gestures. There was a main effect of the condition, $\chi^2(2) = 214.91$, $p < .001$. For both spatial relation types, CG condition was associated with an increased proportion of fixations to the gestural space by $.21 \pm .01$ compared to SO condition ($z = 14.32$, $p < .001$) and by $.06 \pm .01$ compared to RG condition, $z = 4.45$, $p < .001$. The proportion of fixations to gestural space was also higher in the RG condition compared to SO condition by $.14 \pm .01$, $z = 9.87$, $p < .001$. There was also a main effect of the spatial relation type, $\chi^2(1) = 4.47$, $p = .03$. Across all conditions, on–under trials were associated with increased proportions of fixations to the gestural space by $.02 \pm .01$ compared to left–right trials, $z = 2.12$, $p = .03$. There was no interaction between condition and spatial relation type, $\chi^2(2) = 1.03$, $p = .60$.

In sum, participants fixated gestures more in the CG condition compared to the RG condition. Also, they fixated on the gestural space for on–under trials more compared to left–right trials.

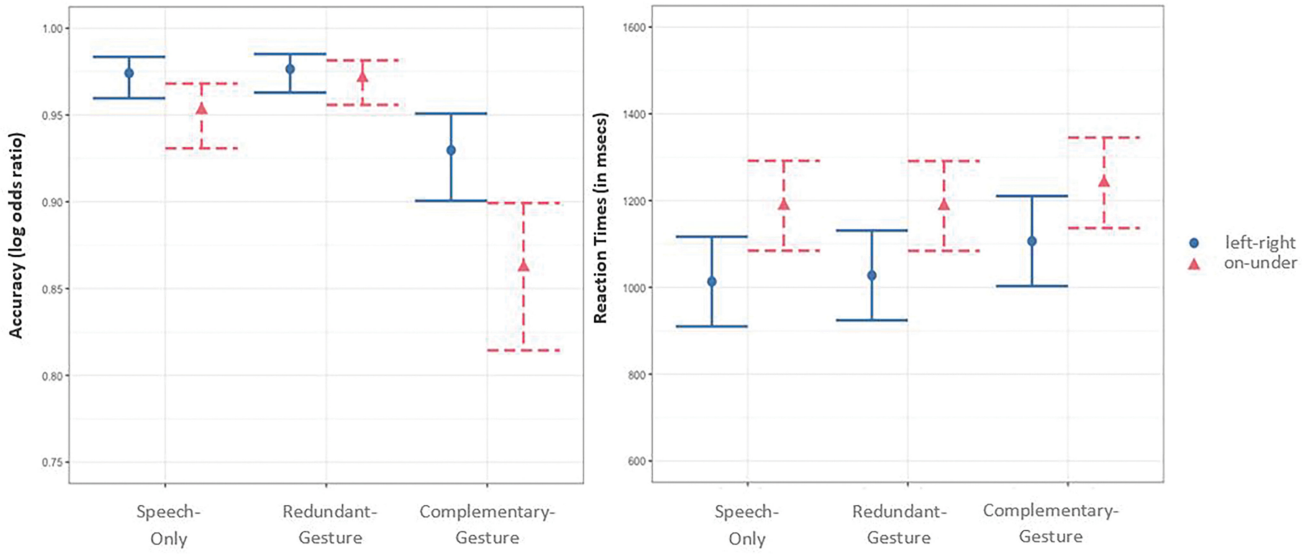
Eye-Tracking and Experimental Task Performance Results

Last, we asked how gazing at gestures during the second sentence related to task performance (accuracy in Model 4 and RT in Model 5) across different conditions and spatial relation types. The fixed effects in these models included the main effects of the condition, the spatial relation type, the proportion of fixations to the gestural

[1] We also analyzed the proportion of fixations to gestural space during the first sentence (i.e., the introductory sentence that was equal across trials and during which there were no gestures). Results showed that fixations to gestural space during the first sentence did not differ across conditions, $\chi^2(2) = 0.75$, $p = .69$, and spatial relation types, $\chi^2(1) = 0.41$, $p = .52$. There was also no interaction between condition and spatial relation type, $\chi^2(5) = 4.52$, $p = .48$.

Figure 4

Accuracy (Left) and RTs (Right) Across Different Conditions and Spatial Relation Types



Note. The brackets represent 95% confidence intervals. RTs = reaction times. See the online article for the color version of this figure.

space, and the two- and three-way interactions between these variables. Please note that we only included the RG and the CG conditions in Models 4 and 5 as there were no gestures in the SO condition.

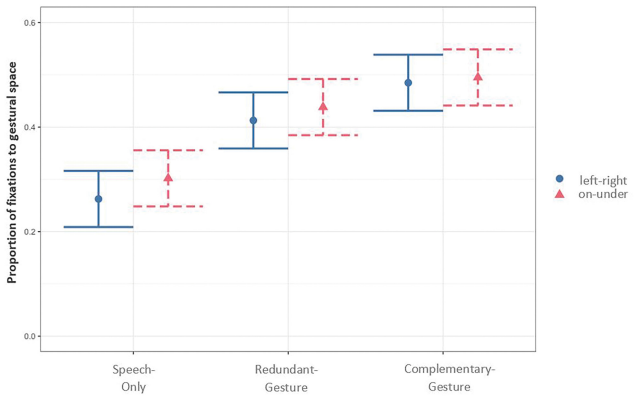
Accuracy results (Model 4, see the top graph in Figure 6) showed no main effect of gazing at gestures, $\chi^2(1) = 0.45$, $p = .50$. There were no two-way interactions between gazing at gestures and condition, $\chi^2(1) = 2.68$, $p = .10$, or between gazing at gestures and spatial relation type, $\chi^2(1) = 2.52$, $p = .11$. There was also no three-way interaction between gazing at gestures, condition, and spatial relation type, $\chi^2(1) = 0.10$, $p = .76$.

RT results (Model 5, see the bottom graph in Figure 6) revealed a main effect of gazing at gestures, $\chi^2(1) = 7.63$, $p = <.001$. Higher proportions of fixations to gestural space were associated with longer RTs by 80.48 $\pm$ 30.64, $t = 2.63$, $p < .01$. There were no two-way interactions between gazing at gestures and condition, $\chi^2(1) = 2.89$, $p = .09$, between gazing at gestures and spatial relation type, $\chi^2(1) = 2.06$, $p = .15$. There was no three-way interaction between gazing at gestures, condition, and spatial relation type, $\chi^2(1) = 1.66$, $p = .20$.

In summary, we found that gazing at gestures did not relate with accuracy. However, fixations to gestures were associated with increased RTs across conditions and spatial relation types.

Figure 5

The Proportion of the Number of Fixations to the Gestural Space AOI Across Different Conditions and Spatial Relation Types



Note. The brackets represent 95% confidence intervals. AOI = areas of interest. See the online article for the color version of this figure.

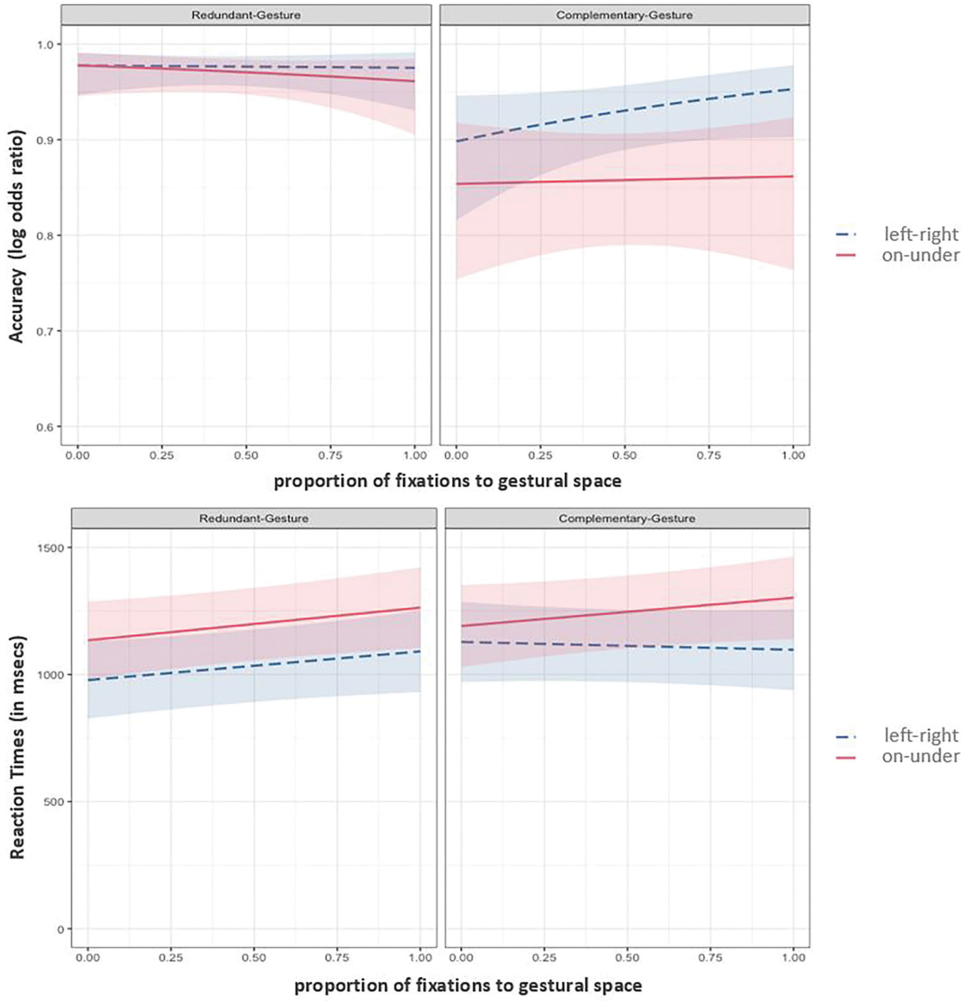
Discussion

This study investigated how listeners attend to and process gestures during spatial language comprehension when gestures have different semantic relations to speech and are communicatively cued differently by speech. Specifically, we asked whether and how (a) listeners gazed at gestures that are co-produced and thus cued to have heightened communicative importance for the comprehension of message by demonstratives in speech and (b) gazing at gestures modulated listeners' comprehension of multimodal message for spatial language. Additionally, we asked whether these measures were modulated by the type of spatial relation: viewpoint-dependent (*left–right*) and viewpoint-independent (*on–under*) spatial relations.

Our results showed that (a) as we predicted, comprehension got hindered when gestures were complementary to speech (i.e., when used along with demonstratives) and more for *on–under* than *left–right*, (b) contrary to our predictions, comprehension did not differ across RG and SO conditions for both spatial relation types, (c) in line with our prediction, listeners fixated gestures more when they complemented demonstratives compared to when expressed redundant information to the accompanying speech for both spatial types, and (d) unlike our prediction, gazing at gestures did not relate

Figure 6

Accuracy (Top) and RTs (Bottom) as a Function of Gesture Gazing Across Different Conditions and Spatial Relation Types



Note. The *x*-axis shows the proportion of fixations to the gestural space. The hues around the lines represent 95% confidence intervals. RTs = reaction times. See the online article for the color version of this figure.

to increased information uptake regardless of the type of spatial relations

Task Performance: How Does Observing Gestures Affect Listeners' Comprehension of Different Types of Spatial Relations?

Contrary to earlier evidence suggesting that observing gestures facilitate listeners' comprehension (Beattie & Shovelton, 1999a, 1999b, 2002a; Holler et al., 2009; Hostetter et al., 2018), our results demonstrated no difference in accuracies or RTs across multimodal (speech + gesture, i.e., RG) and unimodal (i.e., SO) encodings for both types of spatial relations (for null effects, see Goldin-Meadow & Singer, 2003; Kelly & Goldsmith, 2004; Krauss et al., 1995; McNeil et al., 2000). Gestures might be particularly beneficial for comprehension (i.e., compared to SO) in challenging communicative situations (e.g., degraded speech or nonnative comprehension, Drijvers & Özyürek, 2018; Holle et al., 2010; Sueyoshi & Hardison, 2005) or for people with less language proficiency (e.g., children vs. adults, Hostetter, 2011). However, in the current study, we tested verbally competent adults who were native listeners in clear speech comprehension. In addition, average accuracies for RG (96%) and SO (95%) conditions showed a ceiling effects, suggesting that participants might have found the task easy, and there may not be room for gestural enhancement. However, in line with our predictions, we found that participants were less accurate when gestures were the only source of information (when they were used with a demonstrative) for the successful comprehension compared to the other two conditions in which the spatial relation was expressed in the spoken modality (RG and SO). This suggests that participants could extract

categorical spatial information (e.g., leftness) more readily from speech compared to gestures, particularly when gestures do not convey more nuanced categorical spatial relation, such as bottom left.

We predicted that comprehension would be lower for viewpoint-dependent spatial relations (*left–right*) compared to viewpoint-independent spatial relations (*on–under*) as left–right gestures that were executed by the speaker's egocentric perspective require viewpoint alignment between interlocutors (Galati et al., 2013; Hostetter et al., 2018; Karadöller, Sümer, Ünal, & Özyürek, 2021; Keysar et al., 2000). However, as a reverse effect, our results showed that participants had lower accuracies and longer RTs for *on–under* compared to *left–right* across all conditions. There might be several reasons for the current finding. First, participants might find it challenging to distinguish on versus under gestures as they were executed in a smaller and vertical gesture space with subtle differences in hand movements instead of left–right gestures that were executed in a larger and horizontal gesture space and indexed the location more broadly with gross hand/arm movements. Second, figure and ground objects were always in contact in our stimuli set for the on–under spatial relations. In contrast, objects always stand apart from each other in left–right spatial relations. Especially for "under" relations, in most pictures, only a portion of the figure object was visible as the ground object occluded it. This might have hindered participants' performance as they had problems distinguishing objects. Third, people might have certain beliefs on which figure-ground object pairs afford on–under spatial relations. We generally tend to think of the bigger object as the ground when conceptualizing on–under relations with contact. For example, we tend to conceptualize that "the pencil is on the paper" instead of "the paper is under the pen." Although the ground objects were always bigger than figure objects in our stimuli set, participants might still get confused about whether and how figure-ground object pairs afford on–under relations compared to left–right relations devoid of this problem. Future work should pay attention to these stimuli characteristics and the use of space for spatial relations while designing studies.

Eye Gaze: How Do Listeners Fixate Complementary Versus Redundant Gestures for Different Types of Spatial Relations?

Participants looked at the gestural space approximately 25% of the total viewing time when there was a gesture (excluding SO condition, see Table S2). This proportion is higher compared to the previous findings (e.g., 0.5% in Gullberg & Holmqvist, 1999; 8% in Gullberg & Kita, 2009; 9.3% in Yeo & Alibali, 2017). However, it is important to note that this difference between visual attention to gestures might stem from differences in the size of gestural space AOIs, which were operationalized as fixations directly to hands instead of a rather larger space in which gestures are executed (but see Drijvers et al., 2019 for a similar approach). In the current study, we preferred to use a rather larger gestural space for fixations as the gestures that were used in the current paradigm were deictic (i.e., locational) gestures. In such a case, the location information can be extracted by gross hand and arm movements without necessarily requiring information uptake from the hand configuration. Thus, differences across earlier studies and the current study might stem from coding artifacts. Moreover, earlier studies measured eye movements in relatively more natural communication settings such as live interactions (e.g., Gullberg & Holmqvist, 1999) or in videos with speakers selected from the corpus (e.g., Gullberg & Kita, 2009; but see Yeo & Alibali, 2017). In those studies, gestures were not acted as they were in the current paradigm. Rather, gestures in earlier studies were more natural and part of the continuous visual noise, embedded in speech. In the current paradigm, however, gestures were scripted and stood out as visual singletons in the videos, which might attract more visual attention.

Crucially in line with the main prediction of the current study, our results showed that listeners gazed more at gestures that complement speech (with demonstratives) compared to the ones that expressed redundant information to speech. Earlier research showed that when gestures provide a better and an alternative channel of information than speech, listeners allocate more overt visual attention to those gestures, as in the case of communicating in a nonnative language (e.g., Drijvers et al., 2019), when gestures resolve speech ambiguity (e.g., Yeo & Alibali, 2017), or when speech is less informative (van Nispen et al., 2022). Potentially, listeners actively forage alternative sources of disambiguating information when there are certain insufficiencies in speech, which also guides visual attention to gestures. This suggests that gestures might have heightened communicative value, depending on the quality of the speech. However, it is important to demarcate this from other factors such as explicit cues that signal the gesture to be communicative. Listeners might also passively attend to gestures when they are signaled to be central for the successful comprehension of the multimodal package by speaker's explicit cues. One such case is when gestures are co-produced with demonstratives (Cooperrider, 2017). The current study is among the first to investigate visual attention to gestures that are used along with demonstratives. We showed that the heightened communicative value of gestures as signaled by the concurrent use of demonstratives guides listeners' attention to gestures. This also aligns with the evidence that visual cues, such as speakers' fixations to their own gestures (i.e., auto-fixations), also marks the importance and the relevance of speakers' gestures and direct listeners' attention to those gestures (Gullberg & Holmqvist, 1999; Gullberg & Kita, 2009; see Emmorey et al., 2008 for a similar finding in sign language).

We also found that listeners gazed at *on–under* gestures more compared to *left–right* gestures across both redundant- and CG conditions. This might emerge from differences in hand and arm movements across *left–right* and *on–under* gestures discussed above. The *left–right gestures* were executed in a larger space with gross hand/arm movements in an indexical manner to show left versus right space, whereas *on–under* gestures were executed in a smaller space with less gross hand movements. However, although left versus right gestures were executed comparable articulatory spaces in relation to the torso, there were differences in terms of gestural space across on versus under gestures (see Figure 2). That is why, we carried out extra analyses to examine fixations to gestures across left versus right and on versus under gestures, separately (see online supplemental materials). Results showed that participants' fixations to the gestural space did not differ across left versus right gestures for both the redundant- and the CG conditions. However, participants fixated gestural space more for "under" gestures that were executed in a lower peripheral space than "on" gestures both for the redundant- and CG conditions. This finding aligns with earlier evidence showing that gestures that were executed in the peripheral space (i.e., away from the body and the face) might attract more attention, particularly for the gestures

that are produced in the vertical axis (Gullberg & Holmqvist, 1999). Thus, differences across left–right versus on–under gestures might particularly stem from "under" gestures that were executed in a lower peripheral space compared to other gestures.

Eye Gaze and Comprehension: Does Gazing at Gestures Relate to How Much Listeners Benefit From Observing Gestures?

We predicted that gazing at gestures would be related to enhanced comprehension. However, our results showed no relation between gazing at gestures and comprehension for accuracy measure (see Beattie et al., 2010; Gullberg & Kita, 2009; but see Drijvers et al., 2019, for null effects). This suggests that in the current paradigm, the information conveyed through gestures can be extracted peripherally without necessarily fixating to them. This finding aligns with our eye-gaze finding that listeners might "obediently" attend to those gestures as they are cued by an explicit signal (i.e., demonstrative) with no apparent benefit for comprehension (Cooperrider, 2017).

Although we did not find any relation in accuracy, our results suggested that overall fixations to gestures were related to slower RTs. We believe that this effect might be due to problems differentially related to on–under versus left–right gestures that were used in the current paradigm. First for on–under gestures, we showed that although participants fixated gestures more when they complemented speech, they had lower accuracy in that condition. This shows that although participants fixate gestures more when they are cued to be informative by speech (i.e., in the CG condition), gestures might require speech to be interpreted. This is particularly the case for on–under gestures. Participants had harder time to distinguish on–under gestures that were executed in a smaller space in the vertical axis. On the other hand, left–right gestures indexed location in a relatively wider and distinguishable gestural space, considering the body as the ground/relatum. Second, complementing this, our results showed that participants were slower to respond as they fixated gestures more across both redundant- and CG conditions for both left–right and on–under gestures. We believe this might stem from the fact that locational gestures might require extra cognitive processing, which eventually lead to slower responses: the mismatching location indexed by left–right gestures and the relatively indistinguishable on–under gestures might create extra cognitive burden for listeners (Hostetter et al., 2018).

One important area that should be investigated in further research is the mechanisms that govern individual differences in visual attention to gestures (see Özer & Göksun, 2020a for review). It is mostly unknown how individual differences in cognitive resources (e.g., spatial and verbal skills, Hostetter & Alibali, 2011; Nagels et al., 2015; Özer & Göksun, 2020b) relates to visual attention to and processing of gestures. Future work could examine the relation between gazing at gestures and information uptake by employing more implicit measures of processing (such as electrophysiological and neuroimaging measures, Özyürek et al., 2007; Willems et al., 2009) or different measures of attention allocation, such as pupil size (Binda et al., 2014).

Conclusion

This study investigated listeners' visual attention to gestures as a function of the information value of gestures, cued by the speech (i.e., demonstratives). We demonstrated that listeners allocated more overt visual attention to gestures when they complemented demonstratives in speech than when they expressed redundant information that has been already present in the speech. However, direct visual attention to gestures did not modulate listeners' comprehension. These results suggest that gestures that are cued to be central by explicit signals, such as demonstratives in speech, guides listeners' visual attention. However, listeners can process gestures peripherally without fixating them as fixating gestures that are cued to be communicative have no apparent benefit for comprehension.

References

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition & Computation*, *5*(4), 307–331. https://doi.org/10.1207/s15427633scc0504_2

Allen, G. L. (2003). Gestures accompanying verbal route directions: Do they point to a new avenue for examining spatial representations? *Spatial Cognition & Computation*, *3*(4), 259–268. https://doi.org/10.1207/s15427633scc0304_1

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge University Press.

Argyle, M., & Graham, J. A. (1976). The central Europe experiment: Looking at persons and looking at objects. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 6–16. https://doi.org/10.1007/BF01115461

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). *Fitting linear mixed-effects models using lme4*. http://arxiv.org/abs/1406.5823

Beattie, G., & Shovelton, H. (1999a). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, *123*(1–2), 1–30. https://doi.org/10.1515/semi.1999.123.1-2.1

Beattie, G., & Shovelton, H. (1999b). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology*, *18*(4), 438–462. https://doi.org/10.1177/0261927X99018004005

Beattie, G., & Shovelton, H. (2002a). An experimental investigation of the role of different types of iconic gesture in communication: A semantic feature approach. *Gesture*, *1*(2), 129–149. https://doi.org/10.1075/gest.1.2.03bea

Beattie, G., & Shovelton, H. (2002b). What properties of talk are associated with the generation of spontaneous iconic hand gestures? *British Journal of Social Psychology*, *41*(3), 403–417. https://doi.org/10.1348/014466602760344287

Beattie, G., & Shovelton, H. (2006). When size really matters: How a single semantic feature is represented in the speech and gesture modalities. *Gesture*, *6*(1), 63–84. https://doi.org/10.1075/gest.6.1.04bea

Beattie, G., Webster, K., & Ross, J. (2010). The fixation and processing of the iconic gestures that accompany talk. *Journal of Language and Social Psychology*, *29*(2), 194–213. https://doi.org/10.1177/0261927X09359589

Binda, P., Pereverzeva, M., & Murray, S. O. (2014). Pupil size reflects the focus of feature-based attention. *Journal of Neurophysiology*, *112*(12), 3046–3052. https://doi.org/10.1152/jn.00502.2014

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, *137*(4), 706–723. https://doi.org/10.1037/a0013157

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116. https://doi.org/10.1037/a0021790

Cooperrider, K. (2017). Foreground gesture, background gesture. *Gesture*, *16*(2), 176–202. https://doi.org/10.1075/gest.16.2.02coo

Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension.

Psychological Bulletin, *145*(8), 765–784. https://doi.org/10.1037/bul0000202

de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science*, *4*(2), 232–248. https://doi.org/10.1111/j.1756-8765.2012.01183.x

Drijvers, L., & Özyürek, A. (2017). Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *Journal of Speech, Language, and Hearing Research*, *60*(1), 212–222. https://doi.org/10.1044/2016_JSLHR-H-16-0101

Drijvers, L., & Özyürek, A. (2018). Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions. *Brain and Language*, *177–178*, 7–17. https://doi.org/10.1016/j.bandl.2018.01.003

Drijvers, L., Vaitonytė, J., & Özyürek, A. (2019). Degree of language experience modulates visual attention to visible speech and iconic gestures during clear and degraded speech comprehension. *Cognitive Science*, *43*(10), Article e12789. https://doi.org/10.1111/cogs.12789

Driskell, J. E., & Radtke, P. H. (2003). The effect of gesture on speech production and comprehension. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(3), 445–454. https://doi.org/10.1518/hfes.45.3.445.27258

Emmorey, K., & Casey, S. (2001). Gesture, thought and spatial language. *Gesture*, *1*(1), 35–50. https://doi.org/10.1075/gest.1.1.04emm

Emmorey, K., Thompson, R., & Colvin, R. (2008). Eye gaze during comprehension of American Sign Language by native and beginning signers. *Journal of Deaf Studies and Deaf Education*, *14*(2), 237–243. https://doi.org/10.1093/deafed/enn037

Emmorey, K., Tversky, B., & Taylor, H. A. (2000). Using space to describe space: Perspective in speech, sign, and gesture. *Spatial Cognition and Computation*, *2*(3), 157–180. https://doi.org/10.1023/A:1013118114571

Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge University Press.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.

Galati, A., & Avraamides, M. N. (2013). Flexible spatial perspective-taking: Conversational partners weigh multiple cues in collaborative tasks. *Frontiers in Human Neuroscience*, *7*, Article 618. https://doi.org/10.3389/fnhum.2013.00618

Galati, A., Michael, C., Mello, C., Greenauer, N. M., & Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *Journal of Memory and Language*, *68*(2), 140–159. https://doi.org/10.1016/j.jml.2012.10.001

Gerwing, J., & Allison, M. (2009). The relationship between verbal and gestural contributions in conversation: A comparison of three methods. *Gesture*, *9*(3), 312–336. https://doi.org/10.1075/gest.9.3.03ger

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, *136*, 359–364. https://doi.org/10.1016/j.cognition.2014.11.040

Göksun, T., Goldin-Meadow, S., Newcombe, N., & Shipley, T. (2013). Individual differences in mental rotation: What does gesture tell us? *Cognitive Processing*, *14*(2), 153–162. https://doi.org/10.1007/s10339-013-0549-1

Göksun, T., Lehet, M., Malykhina, K., & Chatterjee, A. (2013). Naming and gesturing spatial relations: Evidence from focal brain-injured individuals. *Neuropsychologia*, *51*(8), 1518–1527. https://doi.org/10.1016/j.neuropsychologia.2013.05.006

Goldin-Meadow, S. (2003). *How our hands help us think*. Harvard University Press.

Goldin-Meadow, S., & Singer, M. A. (2003). From children's hands to adults' ears: Gesture's role in the learning process. *Developmental Psychology*, *39*(3), 509–520. https://doi.org/10.1037/0012-1649.39.3.509

Gullberg, M., & Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics & Cognition*, *7*(1), 35–63. https://doi.org/10.1075/pc.7.1.04gul

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition*, *14*(1), 53–82. https://doi.org/10.1075/pc.14.1.05gul

Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior*, *33*(4), 251–277. https://doi.org/10.1007/s10919-009-0073-2

Holle, H., Obleser, J., Rueschemeyer, S. A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage*, *49*(1), 875–884. https://doi.org/10.1016/j.neuroimage.2009.08.058

Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior*, *33*(2), 73–88. https://doi.org/10.1007/s10919-008-0063-9

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology*, *26*(1), 4–27. https://doi.org/10.1177/0261927X06296428

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, *137*(2), 297–315. https://doi.org/10.1037/a0022128

Hostetter, A. B., & Alibali, M. W. (2011). Cognitive skills and gesture–speech redundancy: Formulation difficulty or communicative strategy? *Gesture*, *11*(1), 40–60. https://doi.org/10.1075/gest.11.1.03hos

Hostetter, A. B., Alibali, M. W., & Bartholomew, A. (2011). Gesture during mental rotation. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 1448–1453). Cognitive Science Society.

Hostetter, A. B., Murch, S. H., Rothschild, L., & Gillard, C. S. (2018). Does seeing gesture lighten or increase the load? Effects of processing gesture on verbal and visuospatial cognitive load. *Gesture*, *17*(2), 268–290. https://doi.org/10.1075/gest.17017.hos

Karadöller, D. Z. (2022). *Development of spatial language and memory: Effects of language modality and late sign language exposure* [PhD Thesis]. Radboud University Nijmegen.

Karadöller, D. Z., Sümer, B., & Özyürek, A. (2021). Effects and non-effects of late language exposure on spatial language development: Evidence from deaf adults and children. *Language Learning and Development*, *17*(1), 1–25. https://doi.org/10.1080/15475441.2020.1823846

Karadöller, D. Z., Sümer, B., Ünal, E., & Özyürek, A. (2021). Spatial language use predicts spatial memory of children: Evidence from sign, speech, and speech-plus-gesture. In T. Fitch, C. Lamm, H. Leder, & K. Teßmar-Raible (Eds.), *Proceedings of the 43rd annual conference of the cognitive science society (CogSci 2021)* (pp. 672–678). Cognitive Science Society.

Karadöller, D. Z., Sümer, B., Ünal, E., & Özyürek, A. (2022). Sign advantage: Both children and adults' spatial expressions in sign are more informative than those in speech and gestures combined. *Journal of Child Language*. Advance online publication. https://doi.org/10.1017/S0305000922000642

Karadöller, D. Z., Ünal, E., Sümer, B., Göksun, T., Özer, D., & Özyürek, A. (2019). *Children but not adults use both speech and gesture to produce informative expressions of Left–Right relations*. The 44th annual Boston University conference on language development, Boston, USA.

Kelly, S. D., & Goldsmith, L. H. (2004). Gesture and right hemisphere involvement in evaluating lecture material. *Gesture*, *4*(1), 25–42. https://doi.org/10.1075/gest.4.1.03kel

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension.

Psychological Science, *11*(1), 32–38. https://doi.org/10.1111/1467-9280.00211

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*(3), 396–414. https://doi.org/10.1016/j.jml.2007.06.005

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gesture. *Journal of Experimental Social Psychology*, *31*(6), 533–552. https://doi.org/10.1006/jesp.1995.1024

Lavergne, J., & Kimura, D. (1987). Hand movement asymmetry during speech: No effect of speaking topic. *Neuropsychologia*, *25*(4), 689–693. https://doi.org/10.1016/0028-3932(87)90060-1

Lenth, R. V. (2021). *emmeans: Estimated marginal means, aka least-squares means*. R package (Version 1.6.0) [Computer software]. https://CRAN.R-project.org/package=emmeans

Long, J. A. (2020). *jtools: Analysis and presentation of social scientific data*. R package (Version 2.1.0) [Computer software]. https://cran.r-project.org/package=jtools

McNeil, N. M., Alibali, M. W., & Evans, J. L. (2000). The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior*, *24*(2), 131–150. https://doi.org/10.1023/A:1006657929803

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Melinger, A., & Levelt, W. J. M. (2005). Gesture and the communicative intention of the speaker. *Gesture*, *4*(2), 119–141. https://doi.org/10.1075/gest.4.2.02mel

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. The Belknap Press of Harvard University.

Nagels, A., Kircher, T., Steines, M., Grosvald, M., & Straube, B. (2015). A brief self-rating scale for the assessment of individual differences in gesture perception and production. *Learning and Individual Differences*, *39*, 73–80. https://doi.org/10.1016/j.lindif.2015.03.008

Nobe, S., Hayamizu, S., Hasegawa, O., & Takahashi, H. (1997, September). Are listeners paying attention to the hand gestures of an anthropomorphic agent? An evaluation using a gaze tracking method. In *International gesture workshop* (pp. 49–59). Springer, Berlin.

Nobe, S., Hayamizu, S., Hasegawa, O., & Takahashi, H. (2000). Hand gestures of an anthropomorphic agent: Listeners' eye fixation and comprehension. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, *7*(1), 86–92. https://doi.org/10.11225/jcss.7.86

Özer, D., & Göksun, T. (2020a). Gesture use and processing: A review on individual differences in cognitive resources. *Frontiers in Psychology*, *11*, Article 573555. https://doi.org/10.3389/fpsyg.2020.573555

Özer, D., & Göksun, T. (2020b). Visual-spatial and verbal abilities differentially affect processing of gestural vs. spoken expressions. *Language, Cognition and Neuroscience*, *35*(7), 896–914. https://doi.org/10.1080/23273798.2019.1703016

Özer, D., Tansan, M., Özer, E. E., Malykhina, K., Chatterjee, A., & Göksun, T. (2017). The effects of gesture restriction on spatial language in young and elderly adults. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 1471–1476). Cognitive Science Society.

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605–616. https://doi.org/10.1162/jocn.2007.19.4.605

Posner, M. I. (2016). Orienting of attention: Then and now. *Quarterly Journal of Experimental Psychology*, *69*(10), 1864–1875. https://doi.org/10.1080/17470218.2014.937446

Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives* (Technical Report DAMTP 2009/NA06). Centre for Mathematical Sciences, University of Cambridge.

Pyers, J. E., Perniss, P., & Emmorey, K. (2015). Viewpoint in the visual-spatial modality: The coordination of spatial perspective. *Spatial Cognition & Computation*, *15*(3), 143–169. https://doi.org/10.1080/13875868.2014.1003933

Rimé, B., Boulanger, B. & d'Ydewalle, G. (1988, August 28–September 2). *Visual attention to the communicator's nonverbal behavior as a function of the intelligibility of the message* [Paper presentation]. The symposium on TV behavior, 24th international congress of psychology, Sydney, Australia.

RStudio Team. (2020). *RStudio: Integrated development for R. RStudio*. PBC. http://www.rstudio.com/

Slonimska, A., Özyürek, A., & Campisi, E. (2015). Ostensive signals: Markers of communicative relevance of gesture during demonstration to adults and children. In G. Ferre & M. Tutton (Eds.), *Proceedings of the 4th GESPIN—gesture & speech in interaction conference* (pp. 217–222). Universite of Nantes.

So, W.-C., Shum, P. L.-C., & Wong, M. K.-Y. (2015). Gesture is more effective than spatial language in encoding spatial information. *Quarterly Journal of Experimental Psychology*, *68*(12), 2384–2401. https://doi.org/10.1080/17470218.2015.1015431

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*(4), 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

van Nispen, K., Sekine, K., van der Meulen, I., & Preisig, B. C. (2022). Gesture in the eye of the beholder: An eye-tracking study on factors determining the attention for gestures produced by people with aphasia. *Neuropsychologia*, *174*, Article 108315. https://doi.org/10.1016/j.neuropsychologia.2022.108315

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org

Willems, R. M., Özyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage*, *47*(4), 1992–2004. https://doi.org/10.1016/j.neuroimage.2009.05.066

Yeo, A., & Alibali, M. W. (2017). Evidence for overt visual attention to hand gestures as a function of redundancy and speech disfluency. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 38th annual conference of the cognitive science society*. Cognitive Science Society.

Received September 17, 2021
Revision received February 7, 2023
Accepted February 9, 2023 ■