Title
Speech Cues Influence the Interpretation of Metaphoric Gestures

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Pissani, Laura
Demberg, Vera

Publication Date
2026

Speech Cues Influence the Interpretation of Metaphoric Gestures

Laura Pissani (laura.pissani@uni-saarland.de)[1] **& Vera Demberg**[1,2]
[1]Department of Language Science and Technology, Saarland University
[2]Department of Computer Science, Saarland University

Abstract

Language and gesture are argued to form an integrated system where each modality contributes to the meaning of utterances. While gestures are known to influence language comprehension, less is known about how language affects the interpretation of gestures. We investigated whether the interpretation of metaphoric gestures is modulated by multimodal cues. Participants viewed hand gestures depicting the figurative interpretation of metaphors such as *Men are oaks*, where the actor flexed both arms to convey robustness while speaking. Gestures were presented with no additional cues, only lip movements, or audible speech. Participants then rated how well each gesture matched the gestured interpretation (e.g., *robust*), an alternative interpretation related to the metaphor but not to the gesture (e.g., *tall*), and an unrelated distractor (e.g., *hikers*). We found that ratings for alternative interpretations increased when gestures were paired with speech cues, compared to conditions without speech cues, suggesting that gesture comprehension is influenced by verbal content.

Keywords: speech-gesture integration; metaphoric gestures; gesture comprehension

Introduction

Language and gesture are argued to form an integrated system, in which each modality contributes to the underlying meaning of utterances (McNeill, 1992). As McNeill (2005) puts it, gestures "are integrated with linguistic content into growth points and appear, to the speaker, as an unbroken package of semiosis with it." The notion of an *unbroken package* implies a bidirectional relationship between speech and gesture, in which gestures shape the meaning of words and words shape the meaning of gestures.

Most research has focused on the role of gestures in language comprehension, reporting their benefits for disambiguating meaning (Debreslioska & Gullberg, 2020; Holle & Gunter, 2007), conveying additional semantic information (Holler et al., 2009), deepening understanding (Kang & Tversky, 2016; Kang et al., 2013), improving memory performance (Iani & Bucciarelli, 2017; Straube et al., 2009), and predicting upcoming words (Reimer et al., 2024; Ter Bekke et al., 2020). However, relatively few studies have examined how language influences gesture comprehension. The present study focuses on the role of language in *gesture* comprehension by examining whether the interpretation of gestures is modulated by additional multimodal cues, such as lip movements and audible speech, compared to when these cues are absent.

From a pragmatic perspective, two views have been offered on how speech and gesture contribute to deriving meaning that depends on the integration of pragmatic knowledge (Kelly et al., 1999). The *additive* contribution hypothesis, attributed to the standard pragmatic view,[1] implies that speech and gesture are interpreted independently and combined only at a later stage. In contrast, the *interactive* contribution hypothesis holds that speech and gesture are processed interactively, with speech providing context for gesture and gesture providing context for speech. Accordingly, the additive view predicts that the interpretation of gestures should be independent of the accompanying speech, whereas the interactive view predicts that the interpretation of speech and gesture influence one another. For example, the expression *Lawyers are sharks* may be understood differently depending on the gestures that accompany it—or on the absence of gestures—and, conversely, the interpretation of a finger tap to the temple may vary depending on whether it is paired with speech or other multimodal cues.

Consistent with the interactive contribution hypothesis, Kelly et al. (1999) observed that the interpretation of pointing gestures was modulated by the presence of speech. Participants viewed videos in which an actor produced an indirect request (e.g., *This isn't a twist-off*) while pointing to an object (e.g., a bottle opener). In the gesture-only condition, participants saw the gesture without speech, whereas in the gesture-speech condition, both modalities were available. Accuracy in identifying the referent was higher when speech accompanied the gesture, suggesting that gesture comprehension is at least partially modulated by speech. As the authors note, a pointing gesture toward the bottle opener could also indicate the table on which it was placed or another nearby object, with speech constraining the set of plausible referents.

Deictic gestures such as pointing serve to direct attention to a referent (Cochet & Vauclair, 2014) or to highlight aspects of the environment (Liszkowski, 2010), yet a single gesture can indicate multiple potential referents. As such, deictic gestures lack inherent meaning and may rely more on accompanying speech to constrain their interpretation. Extending beyond this, we investigated whether the interpretation of *metaphoric*

[1]The standard pragmatic view rests on observations and principles from linguistics, psychology, and philosophy (Clark & Lucy, 1975; Grice, 1975; Searle, 1979). It assumes that the conveyed meaning is derived only after the literal meaning is computed, evaluated against the context, and rejected if it does not fit (for a detailed discussion, see Genovesi, 2020, 2023).

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

gestures is similarly influenced by speech cues.

In the present study, the speech portion consisted of metaphorical expressions of the form *X is Y*, in which the vehicle *Y* predicates properties of the topic *X*. For example, in the expression *Pets are kids*, several properties of kids can be attributed to pets, such as being loving, playful, dependent, and innocent. The gesture accompanying each expression visually depicted one of these properties (e.g., both hands held together at chest level to convey love). Since these gestures represent the figurative content of the metaphor, namely that pets are loving, rather than depicting the concrete referent *kids*, they are metaphoric gestures, as they convey an abstract concept (McNeill, 1992). It is important to note that some degree of iconic resemblance may nevertheless arise (see Cornejo et al., 2009). For example, in *Roads are snakes*, where the property predicated is winding, a hand moving in a curving motion may resemble the shape of a snake. However, because the gesture represents the metaphorical construal of the road rather than a concrete snake, it is still considered metaphoric rather than iconic (see Khatin-Zadeh et al., 2022).

The Current Study

Metaphorical expressions afford multiple interpretations; for example, *Men are oaks* may be interpreted as men being robust or as men being tall. We used the polysemous nature of metaphors to investigate how speech cues influence gesture comprehension by presenting videos of a speaker uttering a metaphorical expression while producing a gesture that supports one interpretation (e.g., saying *Men are oaks* while flexing both arms to convey robustness) across three multimodal conditions (no additional cues, lip movements only, and audible speech). Participants then rated how well each gesture matched the gestured interpretation (*robust*), an alternative interpretation related to the spoken metaphor but not to the gesture (*tall*), and an unrelated interpretation (*hikers*).

We predicted that adding speech cues would increase the perceived match between the gesture and the alternative interpretation. Specifically, when participants see the 'robust' gesture with no speech cues, we expect ratings for the gestured interpretation (e.g., *robust*) to be highest, and ratings for the alternative (e.g., *tall*) and unrelated (e.g., *hikers*) interpretations to be lowest. However, when speech cues become audible (e.g., participants hear *Men are oaks*), we expect ratings for the alternative interpretation (e.g., *tall*) to increase, as men being tall is another plausible interpretation of the metaphor.

We also expected that adding speech cues would slightly increase the perceived match between the gesture and the gestured interpretation, as the speech is supportive of the gesture. In contrast, we expected that ratings for the unrelated interpretation would not vary with the addition of speech cues, as they are unsupported by both the gesture and the speech.

Finally, the lip-movement condition was included to explore whether participants extract speech content from visible lip movements in the absence of audible speech.

Methodology

Participants

A total of 150 native English speakers (age 21–70 years, M = 41.98, SD = 12.03; 84 females) were recruited via Prolific. Participants registered for one of three versions of the study, with 50 participants recruited per group. The study lasted approximately 10 minutes, and participants received £2.50. All were born in and currently residing in the United States.

Materials

We selected 24 metaphors of the form *X is Y* (e.g., *Men are oaks*) from Roncero and de Almeida (2015) and Cornejo et al. (2009). The metaphors were chosen to vary in familiarity and to permit at least two plausible interpretations that could be readily represented through gesture.[2]

For each metaphor, we identified two plausible interpretations and recorded two versions of the same spoken metaphor, each paired with a different gesture. For example, the metaphor *Men are oaks* can be interpreted as men being robust or tall; therefore, we recorded two video clips. In the first clip, the actor flexed both arms to convey robustness, and in the second clip, the actor raised one hand above the head to indicate tall stature.[3] To examine the effect of multimodal cues, we created three versions of each clip. In the baseline condition with no additional cues, participants watched the clip with muted audio and blurred lips (Figure 1a). In the lip-movement condition, participants watched the clip with muted audio and visible lips (Figure 1b). In the speech condition, participants watched the clip with visible lips and audible speech (Figure 1c). Note that the two gestures were independent of each other, such that the 'robust' gesture did not imply tall stature, and vice versa. Thus, a high rating for the interpretation *robust* after viewing the 'robust' gesture does not imply a high rating for the interpretation *tall*.

After each clip played, three interpretations were presented in uppercase, one below the other in random order, each with a rating scale. The gestured interpretation (e.g., *robust*) corresponded directly to the gesture, the alternative interpretation (e.g., *tall*) reflected another interpretation of the metaphor but was unrelated to the gesture, and the unrelated interpretation (e.g., *hikers*) was unrelated to both the gesture and the speech. Table 1 shows examples used in the rating task—the labels of the gestured and alternative interpretations were swapped according to the gesture produced by the speaker.

Procedure

Participants viewed 24 critical trial videos and rated how well each gesture corresponded to each interpretation (gestured, alternative, unrelated) using a five-point rating scale. The multimodal condition was implemented between groups, such

[2] See Pissani and Demberg (2026) for details on the materials and Pissani (2026) for the video stimuli repository.

[3] Although each metaphor was associated with two interpretations and two corresponding gestures, participants viewed only one gesture per metaphor, as they were divided into two lists.

(a) No additional cues **(b) Lip-movement** **(c) Speech cues**

Figure 1: Video frames showing the same 'robust' gesture across the three multimodal conditions.

Table 1: Example of materials used in the rating task.

Metaphor	Gestured	Alternative	Unrelated
Clouds are cotton	FLUFFY	SOFT	SQUARE
Roads are snakes	WINDING	LONG	POISONOUS
People are arrows	DRIVEN	FAST	HURTFUL
Schools are zoos	CRAZY	SMELLY	INFORMATIVE
Trust is glue	BONDING	STRONG	STICKY

that each participant viewed videos with either no additional cues, only lip movements, or speech cues. The interpretations were implemented within subjects, such that all participants rated all three interpretations.

At the start of the experiment, participants received the following instructions: *In this task, you will watch a series of very short video clips. In each clip, a woman will be gesturing. Your task is to rate how well the gesture represents three given interpretations on a scale from 1 (not at all) to 5 (very much).* Each trial also included a reminder to rate how well the gesture corresponded to each interpretation, ensuring that attention remained focused on the gestures.

We included four attention trials featuring easily recognizable gestures (e.g., a 'heart' gesture paired with the interpretations *heart*, *fast*, and *leave*). In these trials, the gestured interpretation was expected to receive the highest rating and the other two the lowest, to ensure that participants attended to the gestures. A SKIP button was provided for participants to use only if they were unable to view a video.

Statistical Analysis

Analyses were conducted in R version 2026.01.0 (R Core Team, 2024). Data processing and visualization were performed using the `tidyverse` package (Wickham et al., 2019), and Bayesian models were performed using the `brms` package (Bürkner, 2017, 2018, 2021).[4]

We fitted Bayesian cumulative logistic regression models to predict the probabilities of ratings moving from one category to the next (e.g., from 1 to 2, from 2 to 3), expressed in log-odds. The full model included weakly informative $N(0, 1)$ priors for the main effects of multimodal cues (no additional cues, lip-movement, speech cues) and interpretation (gestured, alternative, unrelated), as well as their interaction. The reduced model used the same priors but excluded the interaction. Random effects included random intercepts for participants and items. All models converged well ($\hat{R} = 1.00$), and model diagnosis and posterior predictive checks indicated that the models fit the data well.

Results

This section reports descriptive statistics for the rating task, followed by regression models examining the effects of multimodal cues (no additional cues, lip-movement, speech cues) and interpretation (gestured, alternative, unrelated) on ratings of the perceived match between gesture and interpretation.

Descriptive statistics

In the baseline condition, gestured interpretations received higher ratings (M = 3.87, SD = 1.34) than both alternative interpretations (M = 1.96, SD = 1.24) and unrelated interpretations (M = 1.65, SD = 1.07). These results indicate that the gestures were readily understandable even in the absence of additional multimodal cues.

In all attention trials (i.e., gestures of a heart, a thumb, the numbers three, and four), the gestured interpretations received the highest ratings (M = 4.94, SD = 0.40) compared with the alternative (M = 1.07, SD = 0.39) and unrelated interpretations (M = 1.04, SD = 0.31). Therefore, no participants were excluded based on these attention checks, as all rated the gestured interpretations at $M \geq 4.00$ and the alternative or

[4] Data and analysis scripts are available at https://osf.io/dfpzq.

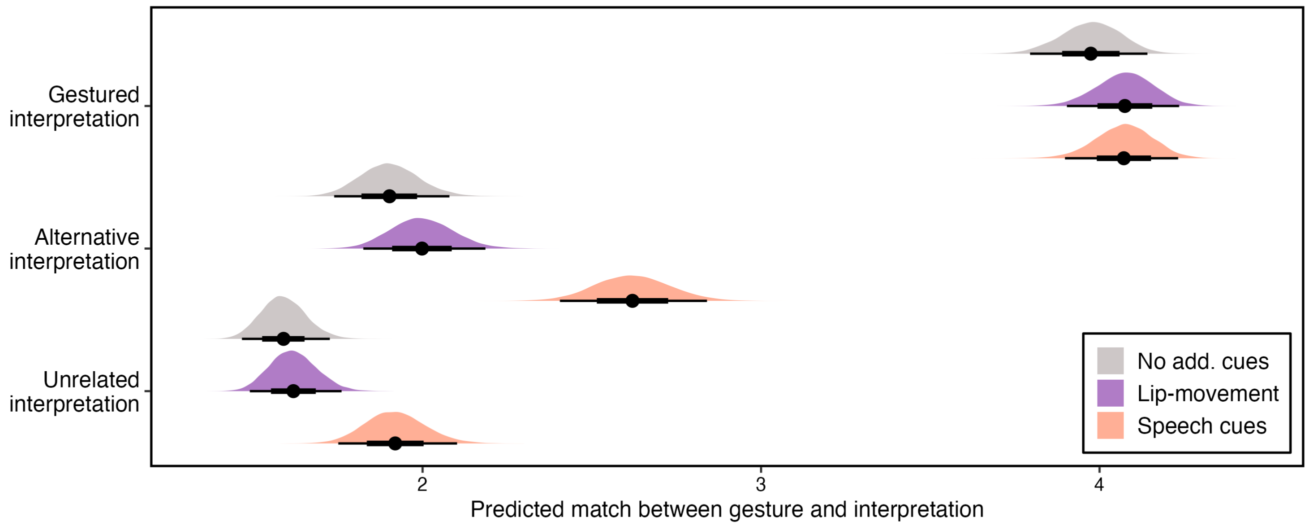

Figure 2: Ratings for the gestured interpretation, alternative interpretation of the metaphor, and unrelated interpretation across multimodal conditions.

unrelated interpretations at $M \leq 2.25$. The SKIP button was used only eight times (0.02% of trials) in the entire study.

Main analysis

This analysis examined whether ratings, measuring how well a gesture matched each interpretation (gestured, alternative, unrelated), varied as a function of multimodal cues (no additional cues, lip-movement, speech cues).

The results of the regression model[5] are summarized in Table 2. We observed a main effect of interpretation, with alternative and unrelated interpretations receiving overall lower ratings than the gestured interpretation. There was an interaction between multimodal cues and interpretation, with Bayes factors[6] providing extreme evidence in favor of the full model including the interaction, $BF_{10} = 10^{16}$.

As Figure 2 shows, the gestured interpretation was more likely to receive the highest ratings, which did not change with the addition of multimodal cues. In contrast, the alternative interpretation—compatible with another interpretation of the spoken metaphor but not with the observed gesture—was more likely to receive higher ratings when presented with speech cues than when presented with no additional cues ($\beta = 0.84$, 95% CRI [0.62, 1.06]), whereas the lip-movement condition did not differ from the condition with no additional cues ($\beta = -0.04$, 95% CRI [−0.26, 0.17]). Interestingly, the unrelated interpretation was also more likely to receive higher ratings when presented with speech cues than when presented with no additional cues ($\beta = 0.41$, 95% CRI [0.19, 0.64]), though to a lesser extent, whereas the lip-movement condition again did not differ from the condition with no additional cues ($\beta = -0.14$, 95% CRI [−0.37, 0.09]).

Table 2: Bayesian regression coefficients for log-odds changes in ratings.

Parameter	$\hat{\beta}$	95% CRI
Fixed effects		
Alternative interpretation	−2.89	[−3.06, −2.73]
Unrelated interpretation	−3.49	[−3.66, −3.32]
Lip-movement	0.19	[−0.13, 0.50]
Speech cues	0.20	[−0.12, 0.51]
Alternative × Lip-movement	−0.04	[−0.26, 0.17]
Unrelated × Lip-movement	−0.14	[−0.37, 0.09]
Alternative × Speech	0.84	[0.62, 1.06]
Unrelated × Speech	0.41	[0.19, 0.64]
Random intercepts		
Item	0.45	[0.33, 0.61]
Participant	0.71	[0.62, 0.80]

Follow-up analyses

The following analyses examined whether the absence of the effect of speech cues for gestured interpretations reflected properties of the metaphors themselves, and whether the unexpected effect of speech cues for unrelated interpretations was driven by the semantic relatedness between those interpretations and individual words in the speech.

[5] Since cumulative logistic regression models predict changes in ratings across response categories rather than the magnitude of the increase, we computed posterior predicted ratings for each interpretation and multimodal condition (Figure 2).

[6] Bayes factors (BF) quantify the strength of evidence for one model over another. We compared a full model (including main effects and the interaction) to a reduced model (excluding the interaction). A BF between 10 and 30 indicates strong evidence, between 30 and 100 indicates very strong evidence, and greater than 100 indicates extreme evidence (Lee & Wagenmakers, 2014).

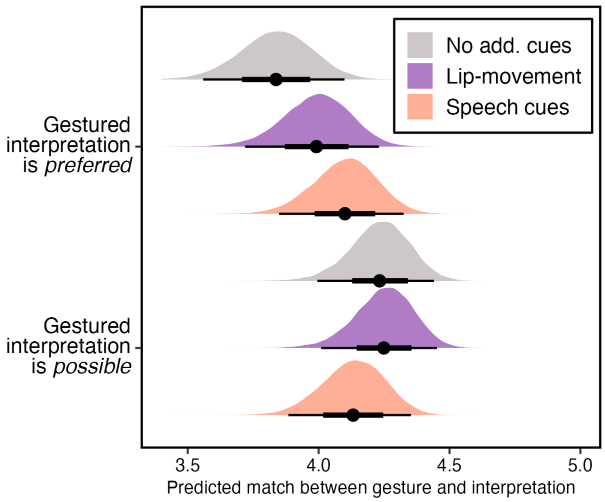

Figure 3: Ratings for gestured interpretations when were preferred versus possible interpretations of the metaphor across multimodal conditions.

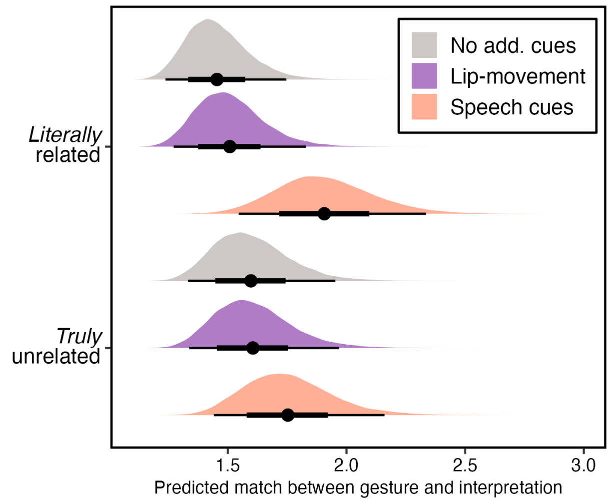

Figure 4: Ratings for unrelated interpretations when they were literally related versus truly unrelated to the metaphor across multimodal conditions.

Gestured Interpretations To investigate why ratings for the gestured interpretation did not differ across multimodal conditions, we examined whether these ratings depended on how metaphors were understood. Since each metaphor had two plausible interpretations—one potentially preferred and the other less preferred but still possible—ratings could vary depending on whether the gestured interpretation was preferred or possible. To assess this, we collected preference ratings using nongestured clips (N = 60). For each metaphor, we calculated the mean rating for each interpretation and coded the one with the highest rating as *preferred* (M = 4.10, SD = 0.43) and the other as *possible* (M = 3.22, SD = 0.57).

We applied the same analytic procedures used in the main analysis to predict ratings for the gestured interpretations as a function of preference, multimodal cues, and their interaction. We observed an interaction between preference and multimodal cues. When the gestured interpretation was the preferred interpretation of the metaphor, it was more likely to receive higher ratings when paired with speech cues compared to no additional cues (β = 0.45, 95% CRI [0.06, 0.84]). In contrast, when the gestured interpretation was the possible interpretation of the metaphor, ratings did not differ between speech cues and no additional cues (β = −0.18, 95% CRI [−0.58, 0.22]). Thus, the effect of speech cues was observed for the gestured interpretation only when it was the preferred interpretation of the metaphor, but not when it was the possible one. In Figure 3, this interaction can be seen by comparing the speech (orange) and no-additional-cues (grey) conditions within each sublevel of the gestured interpretation.

Unrelated Interpretations To investigate why ratings for the unrelated interpretation increased in the speech condition, we examined whether these ratings depended on the semantic relationship between the unrelated interpretation and the individual words in the speech. To this end, we computed cosine similarities between Word2Vec word embeddings[7] for the unrelated interpretation (e.g., *hikers*) and the topic (e.g., *men*) and vehicle (e.g., *oaks*) of each metaphor, separately. Cosine similarity measures the semantic distance between word embeddings, with higher values indicating greater semantic relatedness. Items for which the unrelated interpretation had a similarity greater than 0.25 with either the topic or vehicle were coded as *literally related*, and all others as *truly unrelated*. This procedure yielded 12 literally related interpretations (e.g., *sticky* relative to *glue*) and 12 truly unrelated interpretations (e.g., *square* relative to *cotton*).

We also applied the same analytic procedures used in the main analysis to predict ratings for the unrelated interpretations as a function of relatedness, multimodal cues, and their interaction. We observed an interaction between relatedness and multimodal cues. When the unrelated interpretation was literally related to the individual words in the speech, it was more likely to receive higher ratings when presented with speech cues compared to no additional cues (β = 0.94, 95% CRI [0.49, 1.40]). In contrast, when the unrelated interpretation was truly unrelated to the individual words in the speech, ratings did not differ between speech cues and no additional cues (β = 0.34, 95% CRI [−0.11, 0.79]). Thus, the effect of speech cues was only present for the literally related subset, but not for the truly unrelated subset. In Figure 4, this interaction can be seen by comparing the speech (orange) and no-additional-cues (grey) conditions within each sublevel of the unrelated interpretation.

[7] Word2Vec embeddings (Mikolov, Chen, et al., 2013; Mikolov, Sutskever, et al., 2013) are numerical vector representations of word types learned from a window of surrounding context words, such that words with similar meanings or grammatical roles have similar vectors (Meghdadi et al., 2026).

Discussion

This study examined how gesture comprehension is integrated with multimodal cues by presenting videos of gestures with no additional cues, with only lip movements, and with audible speech. We specifically assessed whether ratings for gestured, alternative, and unrelated interpretations varied with the addition of multimodal cues. The gestured interpretation reflected the interpretation of the metaphor conveyed by the gesture, the alternative interpretation reflected another plausible interpretation of the metaphor in the speech but was unrelated to the gesture, and the unrelated interpretation was independent of both gesture and speech.

As predicted, the perceived match between the gesture and the alternative interpretation increased when the gesture was paired with speech. This pattern is particularly interesting because it shows that participants incorporated the alternative figurative sense of the metaphor even when it was unnecessary for assessing the gesture. When the actor performed the 'robust' gesture while the metaphor *Men are oaks* was audible, ratings for the gestured interpretation (*robust*) remained unchanged, whereas ratings for the alternative interpretation (*tall*) increased compared to when the metaphor was inaudible, despite *tall* being unrelated to the observed gesture. This indicates that participants not only attended to the gesture but also integrated it with both figurative senses of the speech.

In contrast, the perceived match between the gesture and the gestured interpretation did not change when speech cues were added, suggesting that when a gesture is clear, supportive verbal content does not further enhance comprehension. However, when the preferred and possible interpretations of the metaphor were considered, an effect of speech cues emerged when the gestured interpretation was preferred. These results suggest that participants were sensitive to the dominant figurative sense of the metaphor in the speech, which—when aligned with the gesture—contributed to the observed effect.

Surprisingly, the perceived match between the gesture and the unrelated interpretation also increased when presented with speech cues, although to a lesser extent than the alternative interpretation. This may reflect participants' sensitivity to aspects of the speech, including both figurative and literal content. Since these materials were designed to study metaphor comprehension, some unrelated interpretations could be literally associated to the metaphor. For example, as shown in Table 1, the unrelated interpretations *poisonous* and *sticky* could apply literally to a snake and glue, respectively, but not to the figurative senses of *Roads are snakes* or *Trust is glue*. Similarly, *informative* and *hurtful* could describe schools and people, respectively, but are not derived from the figurative senses of *Schools are zoos* or *People are arrows*. To assess this, we divided the unrelated interpretations into literally related and truly unrelated using cosine similarity. Follow-up analyses indicated that the effect of speech cues—previously observed for alternative interpretations—was present for the literally related subset but absent for the truly unrelated subset. This suggests that gesture comprehension was influenced not only by the figurative content of the metaphor but also by the literal meaning of the metaphor in the speech.

In sum, the effect of speech cues was present for *all* interpretations in specific situations: (1) for gestured interpretations, when they matched the dominant figurative sense of the spoken metaphor; (2) for alternative interpretations overall; and (3) for unrelated interpretations, when they were literally related to the spoken metaphor. We did not find an effect of lip-movement information for any interpretation, suggesting that individuals do not reconstruct verbal content from lip movements alone, at least this seems to be the case when the video frame includes the face and torso rather than just a close-up of the face and when no prior context is available to predict subsequent words.

Our findings align with results reported by Kelly et al. (1999), where the interpretation of deictic gestures was modulated by accompanying speech. Although the present study focused on the effects of multimodal cues on the comprehension of metaphoric gestures, our results are consistent with a broader body of work demonstrating the mutual influence of speech and gesture (Cornejo et al., 2009; Kelly, Özyürek, & Maris, 2010; Kelly et al., 1999, 2004; Özer et al., 2026). For example, Cornejo et al. (2009) observed differences in electrophysiological responses between metaphors paired with congruent versus incongruent gestures at early, but not later, processing stages, suggesting that gestures and verbal content are integrated online during metaphor comprehension.

Nevertheless, our results cannot determine whether the integration of speech and gesture occurs automatically, as reported in some studies (e.g., Kelly, Creigh, & Bartolotti, 2010), or under voluntary control, as reported in others (e.g., Holle & Gunter, 2007). Given that both the literal meaning and figurative content of the metaphor influenced gesture interpretation, we consider two possibilities. First, because the literal meaning of metaphors is accessed before the figurative content (Pissani & de Almeida, 2022; Pissani et al., 2026; Weiland et al., 2014) and may linger for about three seconds (Pissani & de Almeida, 2023), concepts associated with the literal meaning could influence gesture interpretation, consistent with a more automatic process. However, given that our videos averaged about five seconds in duration ($M = 5.23$, $SD = 0.57$), it seems less likely that literal concepts would persist while participants were making their judgments. Alternatively, individuals may actively seek meaning even when it is not immediately evident, deliberately deriving the literal meaning of the metaphorical vehicle (e.g., *oaks* in *Men are oaks*), consistent with a more controlled process.

Altogether, these results provide evidence supporting the integrated-systems hypothesis (Kelly, Özyürek, & Maris, 2010), in which speech and gesture are mutually and obligatorily integrated, allowing each modality to influence the other. In the present study, gestures presented without multimodal cues were already interpretable. However, when speech cues were added, additional interpretations of the gesture, consistent with the verbal content, were considered.

Acknowledgments

This research was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 948878).

We thank Sophia Wiedmann for serving as the actor in the video stimuli, Niranjana Satyanarayana for assistance in programming all versions of the experiment, Yuan Zhang for assistance with video editing, and Maryam Meghdadi for assistance computing the cosine similarities.

References

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, *100*(5), 1–54. https://doi.org/10.18637/jss.v100.i05

Clark, H. H., & Lucy, P. (1975). Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, *14*(1), 56–72.

Cochet, H., & Vauclair, J. (2014). Deictic gestures and symbolic gestures produced by adults in an experimental context: Hand shapes and hand preferences. *Laterality: Asymmetries of Body, Brain and Cognition*, *19*(3), 278–301.

Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., López, V., & Núñez, R. E. (2009). Gesture and metaphor comprehension: Electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain and cognition*, *70*(1), 42–52.

Debreslioska, S., & Gullberg, M. (2020). What's new? gestures accompany inferable rather than brand-new referents in discourse. *Frontiers in Psychology*, *11*, 1935.

Genovesi, C. (2020). Metaphor and what is meant: Metaphorical content, what is said, and contextualism. *Journal of pragmatics*, *157*, 17–38.

Genovesi, C. (2023). Grice's café: Coffee, cream, and metaphor comprehension. *Frontiers in Communication*, *8*, 1175587. https://doi.org/10.3389/fcomm.2023.1175587

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: Erp evidence. *Journal of cognitive neuroscience*, *19*(7), 1175–1192.

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.

Iani, F., & Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *Journal of Memory and Language*, *96*, 110–121.

Kang, S., Hallman, G. L., Son, L. K., & Black, J. B. (2013). The different benefits from different gestures in understanding a concept. *Journal of Science Education and Technology*, 825–837.

Kang, S., & Tversky, B. (2016). From hands to minds: Gestures promote understanding. *Cognitive Research: Principles and Implications*, *1*(1), 4.

Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of memory and Language*, *40*(4), 577–592.

Kelly, S. D., Creigh, P., & Bartolotti, J. (2010). Integrating speech and iconic gestures in a stroop-like task: Evidence for automatic processing. *Journal of Cognitive Neuroscience*, *22*, 683–694. https://doi.org/10.1162/jocn.2009.21254

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and language*, *89*(1), 253–260.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological science*, *21*(2), 260–267.

Khatin-Zadeh, O., Farsani, D., & Banaruee, H. (2022). A study of the use of iconic and metaphoric gestures with motion-based, static space-based, static object-based, and static event-based statements. *Behavioral Sciences*, *12*(7), 239.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.

Liszkowski, U. (2010). Deictic and other gestures in infancy. *Acción psicológica*, *7*(2), 21–33.

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

McNeill, D. (2005). *Gesture and thought*. University of Chicago press.

Meghdadi, M., Duff, J., & Demberg, V. (2026). Integrating language model embeddings into the act-r cognitive modeling framework. *Frontiers in Language Sciences*, *5*, 1721326.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, *26*.

Özer, D., Soyman, E., Badakul, A. N., Arslan, B., Yılmaz, F. S., & Göksun, T. (2026). Distinct temporal dynamics of speech and gesture processing: Insights from event-related potentials across l1 and l2. *Journal of Experimental Psychology: General*.

Pissani, L. (2026). *Repository of gestures and multimodal cues*. OSF. https://doi.org/10.17605/OSF.IO/8BR5J

Pissani, L., & de Almeida, R. G. (2022). Can you mend a broken heart? awakening conventional metaphors in the maze.

Psychonomic Bulletin & Review, *29*(1), 253–261. https://doi.org/10.3758/s13423-021-01985-y

Pissani, L., & de Almeida, R. G. (2023). Early birds can fly: Awakening the literal meaning of conventional metaphors further downstream. *Metaphor and Symbol*, *38*(4), 346–362. https://doi.org/10.1080/10926488.2023.2181783

Pissani, L., & Demberg, V. (2026). Bidirectional effects of speech and gesture: Evidence from a novel levels-of-congruence paradigm. [Retrieved from https://osf.io/preprints/psyarxiv/pgfmb_v1]. *PsyArXiv preprint*.

Pissani, L., Scholman, M. C., & Demberg, V. (2026). Working memory and vocabulary knowledge sustain activation of the literal meaning of metaphors during sentence comprehension [Retrieved from https://osf.io/preprints/psyarxiv/dqanh_v1]. *PsyArXiv preprint*.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Reimer, L., Spychalska, M., & Werning, M. (2024). Integratitng co-speech gestures into sentence meaning comprehension. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*.

Roncero, C., & de Almeida, R. G. (2015). Semantic properties, aptness, familiarity, conventionality, and interpretive diversity scores for 84 metaphors and similes. *Behavior research methods*, *47*, 800–812.

Searle, J. R. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 83–111). Cambridge University Press.

Straube, B., Green, A., Weis, S., Chatterjee, A., & Kircher, T. (2009). Memory effects of speech and gesture binding: Cortical and hippocampal activation in relation to subsequent memory performance. *Journal of cognitive neuroscience*, *21*(4), 821–836.

Ter Bekke, M., Drijvers, L., & Holler, J. (2020). The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech. *Gesture and Speech in Interaction Conference*.

Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: Evidence from masked priming erp. *Frontiers in Human Neuroscience*, *8*, 583.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686