Title

Co-speech gestures complement motion state information expressed by verbs

Permalink
https://escholarship.org/uc/item/5r29t582

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Kimura, Hina
Yasuda, Tetsuya
Kobayashi, Harumi

Publication Date
2024

Peer reviewed

Co-speech gestures complement motion state information expressed by verbs

Hina Kimura (23rmd14@ms.dendai.ac.jp)
Graduate School of Tokyo Denki University, Ishizaka, Hatoyama-machi, Hiki-gun
Saitama 350-0394, Japan

Tetsuya Yasuda (t-yasuda@g.ecc.u-tokyo.ac.jp)
The University of Tokyo, 3-8-1, Komaba, Meguro-ku,
Tokyo 153-8902, Japan

Harumi Kobayashi (h-koba@mail.dendai.ac.jp)
Tokyo Denki University, Ishizaka, Hatoyama-machi, Hiki-gun
Saitama 350-0394, Japan

Abstract

Verbs in progressive aspect can be used for different motion phases of people or objects. For example, "A cat is falling" can describe either the beginning of, or on-going, or the ending of the falling of the cat. Then do people spontaneously use different co-speech gestures according to different motion phases when they use the same progressive verb in speech? This study investigated Japanese speakers' co-speech gestures used with a progressive verb in Japanese (verb + progressive morpheme *-teiru*), focusing on the paths of produced gestures. The paths were analyzed according to the direction (vertical or horizontal) or trajectory (arc or straight). The results showed that the participants' use of co-speech gestures differed when they expressed different motion phases (beginning or ending). The study suggests that gestures can compensate the motion phases of agents that may not be described by language.

Keywords: spontaneous gesture, motion phase, gesture path, verbs in linguistic structure

Introduction

When people talk, they often use gestures. Previous studies examined the possible roles of co-speech gestures and showed that gestures can effectively emphasize parts of speech (Bull & Connelly, 1985), enhance listeners' understanding (Kelly, et al., 1999), and help speakers' own thinking and judgment (Goldin-Meadow & Beilock, 2010). One important role is to compensate for lacking information of speech (Iverson & Goldin-Meadow, 2005; Kita & Özyürek, 2003). As for compensation, co-speech gesture may disambiguate ambiguous linguistic structures (Kashiwadate et al., 2020).

In this paper, we wish to add another aspect of compensation for lucking information. We propose that gestures can express the information on motion phases that are not linguistically described. Here, we define that motion phases refer to the beginning, on-going, or the ending (or close to the end) of a motion event.

Verbs are words that typically express human or animal actions or objects' motions. In speech, if a word and the associated iconic gesture are congruent, people better understand and memorize the described content than if these are not, showing the close relationship between speech and gesture and gesture information (Kelly, et al., 2010). Verb aspects such as progressive (e.g., running) or perfective (e.g., finished) are used to express temporal contours of verb meanings. In spontaneous speech, verb meanings are sometimes compensated by gesture. According to Duncan (2002), "… imperfective-progressive aspect-marked spoken utterances regularly accompanied iconic gestures in which the speaker's hands engaged in some kind of temporally extended, repeating, or 'agitated' movements. (p.183)" Even when the verb itself does not describe these motions, important aspects such as manner and path, are spontaneously expressed by gesture (Kita & Özyürek, 2003). These studies showed the important features of iconic co-speech gestures that accompany verbs, but the relationship between motion phases and verbs have not been explored yet.

Although motions phases are important, they are not typically coded with words. For example, the expression "The cat is falling" seems to be applicable to at least three phases of the cat's falling event, the beginning, on-going, and ending of the falling. People may say "The cat is falling!" when the cat began falling motion, or the falling motion is close to the end. For each phase, we may use the same word with appropriate grammatical morpheme "-ing," such as "falling," but strictly speaking, with different meanings. In Japanese, progressive aspect is expressed by a grammatical morpheme "-teiru" (Shirai, 2000). A Japanese expression "*ochi-teiru*" (*ochi* (verb. fall) + *teiru* (progressive morpheme. -ing) can be similarly applicable to each of these phases, the beginning, on-going, and the ending. Actually, in Japanese, -teiru can be used for another phase "resultative," meaning that the object is there as a result of falling (Shirai, 2000). Thus, in Japanese, progressive expression must be differentiated from resultative expression.

Of course, for the expression "*Ochiteiru neko no shasin* (The Falling Cat Photo)," we can specify these different phases by adding more elaborate language, such as "The cat is at the beginning phase of falling!" but such expressions do not sound practical (!). However, co-speech gestures may be able to express such different phases in accordance with verbs. For example, a person may express the beginning of the falling motion using a hand gesture with arc shape and path motion such as straight down movement.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Another aspect of co-speech gestures with production of verbs is that gestures may express an agent of the motion, in addition to the motion itself. For example, the expression "A Falling Cat Photo" can have two meanings, "A photo that depicts a falling cat" and "Falling of the photo that depicts a cat." Thus, the agent of falling motion is "photo" in the first meaning, but "cat" in the second meaning. Hand gestures may express the information about agents even when agent information is ambiguous.

Some previous research suggested that participants' produced gesture reflect underlying linguistic structures even when linguistic structures are ambiguous. For example, Kashiwadate et al. (2020) used an ambiguous phrase "Black Tail Big Cat" in Japanese (*kuroi* (black) *shippo-no* (tail + particle) *okina* (big) *neko* (cat)) that can be interpreted into multiple meanings. Produced gestures revealed that the occurrences of gestures synchronized with linguistic structures. They tended to depict "a big cat" using a hand gesture of a big cat, whereas they depicted "a big tail" using a hand gesture of a big tail according to associated phrase structures. Handa et al. (2021) examined how gestures convey hierarchical structures when participants were given prompt phrases and forced to do gestures to describe the meaning of the prompt. Kimura et al. (2023) also examined how participants describe the meaning of prompt phrase but unlike Kashiwadate et al. and Handa et al., participants' gestures were spontaneous. The results of these studies reported that produced gestures possibly accorded to linguistic structures regardless of forced gestures or spontaneous gestures.

The purpose of our study was two-hold: The first research question was: Do people produce co-speech gestures that express the meaning of the verb in accordance with different phases of a motion event? The second research question was: Do people properly express relevant agents of the verb according to associated linguistic structures?

As for the first question, we used two different phases of a motion event. One was Beginning: An agent was at the beginning phase of a motion. For example, an illustration of Beginning depicted a cat just began falling from a cliff. The other was Ending: An agent was at the end phase of a motion. For example, an illustration of Ending depicted a cat that was finishing falling and close to the ground (Figure 1). These two phases were selected because these depict distinct and contrastive phases of a motion event.

As for the second research question, we utilized ambiguous hierarchical structures that can be interpreted with two different agents. For example, we used a phrase "Falling Cat Photo," that can be interpreted with two different agents, cat, (A cat is falling and this event is depicted by a photo) or a photo (A photo of a cat is falling).

In this study, to examine these two research questions, we also controlled verbs. Previous studies (Handa, et al.; Kimura, et al.) suggested that two verbs, "fall" and "fly" are useful to code different types of motion events, in particular, the beginning and the ending of these. "Fall" suggests a distinct sense of vertical directionality influenced by gravity, so that the beginning and the ending is visually distinct. However, "fly" lacks such a specific directional implication. For flying motion, both gravity and horizontal movement have effects and as a result, the flying object will follow a specific trajectory according to the agent. Therefore, the beginning and the ending may not be so distinct comparing with "fall."

We expected to observe that motion phases would be described with movement of hand gestures that depict either beginning or ending. We also expected that hand shape of such gestures would depict the agent of the motion. We presented a linguistic prompt in addition to the associated illustration, and the participant uttered the prompt and spontaneously used gestures. In the analysis, the speaker's gestures were coded, and occurrences of path gestures were examined using a statistical model and a time series analysis.

Method

Participants

Twenty-eight Japanese monolingual students who spoke Japanese as their first language participated in this study (Mean age = 21.96 years; SD = 1.45 years). We excluded six participants' data because they did not use any gestures, and three participants' data because their gestures were not co-speech gestures. We also excluded two participants data, for a technical problem and a procedural mistake respectively. Finally, the data of seventeen participants who spontaneously produced co-speech gestures were taken for analysis. The experiment was conducted in accordance with the university's code of confidentiality and ethical treatment of human subjects.

Conditions and Stimuli

The experimental conditions (Figure 1) consisted of the Structure type (fall-cat structure {LB：left branching}, cat-photo structure {RB：right branching}), Verb (fall, fly), and Motion phase (beginning, ending).

In the condition of Structure type, the Japanese phrases were used, such as "*Ochi-teiru* (Fall + ing) *neko-no* (Cat + particle) *shashin* (Photo)." This three-word phrase consists of Verb (V: verb) + Noun-1 (N1: the first noun) + Noun-2 (N2: the second noun) (Figure 1). In the fall-cat structure, the Verb and Noun-1 are chunked first, then chunked with Noun-2. This phrase can be interpreted as "a photo that depicts a falling cat (the chunk structure is {{falling, cat}, photo})." In contrast, in the cat-photo structure, Noun-1 and Noun-2 are chunked first then chunked with Verb. This phrase can be interpreted as "a falling photo that depicts a cat (the chunk structure is {falling, {cat, photo}})." Verb condition consisted of two levels: "falling" and "flying." The verb of fall has a strong sense of direction, going down with gravity, like dropping something, while the verb of fly doesn't have a strong direction. These two actions were chosen to show different movements. Motion phase condition consisted of two levels: "beginning" and "ending." The motion of the beginning indicated that, in Fall-cat structure, the cat is beginning of falling, or in Cat-photo structure, the photo is beginning of falling.

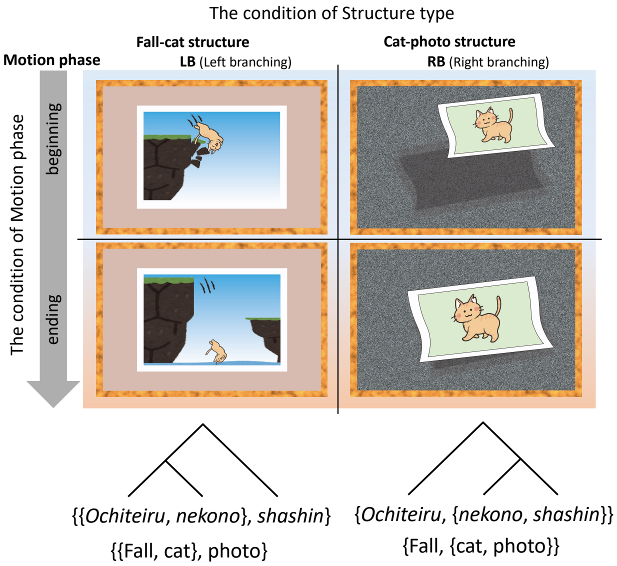

Figure 1: An example of a stimulus "Falling cat photo" which can be interpreted differently. Two Motion phases were used for each structure type: fall-cat structure ("beginning": The cat has begun the process of falling off the cliff. / "ending": The cat is very close to completing the fall.), and cat-photo structure ("beginning": The photo which is depicted the cat has begun to fall. / "ending": The photo is very close to completing its fall.).

The experimental stimuli comprised 12 phrases, each constructed with two verbs (V), six first nouns (N1), and three second nouns (N2). We designed the stimuli to elicit spontaneous gestures, carefully controlling for factors such as a two-mora length and a primary accent on the first mora. For the second nouns, we selected words that represent objects that have similar square shapes, such as photos, stamps, and envelopes, to investigate the expression of the verb more clearly. Twelve pictures were created to represent fall-cat and cat-photo structures with two Motion phases of beginning and ending.

The stimuli comprised two sets, each consisting of a Structure type of animated movies. Each set included 12 trials with pictures representing the conditions under different motion phases. In terms of factorial design, Verb and Motion phase conditions were within-participants factors, whereas the Branching condition was a between-participants factors.

Procedure

The participants were paired up and each participant in a pair was randomly assigned to either the role of the speaker or the listener. The speaker was asked to describe the content of a stimulus. The listener was asked to observe and understand the speaker's expressions about the experimental stimuli.

The speaker and listener sat across at a table facing each other. Each participant was positioned in such a way that they could only see their own monitor and not the other participant's monitor. The experimenter then provided instructions regarding the roles of the speaker and the listener, along with an overview of the experiment's general flow. In addition, the experimenter informed the participants that after the experimental session, the listener will choose one correct illustration from the four illustrations. The participants were also told that they would receive a reward based on the percentage of correct responses after the experiment was completed.

After participants received the instructions, the experimenter asked the speaker look at the specific prompt phrase displayed on the monitor. Then speaker looked at the stimulus and freely conveyed its content to the listener. However, speakers' utterance was restricted only prompt phrase. The listener then answered to a test in which they chose an appropriate picture from a list (a set of four pictures as shown Figure 1). When all the trials were completed, participants switched roles and repeated the entire trial. The participants watched each scene only once because the branching condition was a between-participant factor. That is, no identical pictures were used between the participants in each pair even when their roles were switched.

A digital video camera (FDR-AX40, Sony) was used to record the session. The camera captured the listener's upper body including the arms and the face.

Gesture coding

ELAN (ver. 6.7) was used for analysis. Speech and gestures were coded according to each Structure type based on video data.

We analyzed the gestures based on Kendon's (2004) gesture phases, which capture the movement dynamics of the gesture. The gesture phases included the stroke itself, preparatory movements leading up to the stroke, recovery phase when the gesture withdraws, and post-stroke hold phase when the gesture sustains its position at the end of the stroke. The gesture stroke direction was classified into two categories: vertical and horizontal, because we were interested in whether gestures describe the meanings of the verbs "fall" and "fly," which represent different movements. For "fall" description, vertical movement was expected to be used. For "fly" description, horizontal movement was expected to be used. Quite frequently gesture paths are diagonal but either vertical element or horizontal element seems to be more dominant. Therefore, our classification was based on the path length between the start point and end point of the gesture stroke (Figure 2; this gesture direction was classified as vertical). If the width is relatively longer than the height, then we coded the gesture as "horizontal."

We also classified temporal and spatial information of gestures, distinguishing between arc and straight movements. After we specified the start point and the end point of a stroke on a video frame, we connected these two points and judged whether the line (path) is straight or not straight. All of the gesture strokes that consists of a straight path were coded "straight." All of the gesture strokes that consists of a not-straight path were coded "arc." The start and the end points were appropriately determined on a video frame using a specific point of the participant's hand. In addition, we separately coded pointing gestures in terms of directions and target objects.

We took the onset and end of each uttered words, Verb, Noun-1, and Noun-2., according to the stimulus phrases. The time point at which a word could be heard clearly was identified, and the point was coded as the onset of a word. Then we also identified the end point of the word when the final sound was still heard but difficult to discern, and the point was coded as the end of a word.

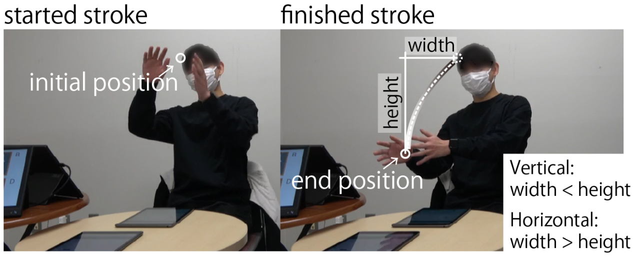

Figure 2: The criterion of "vertical" direction gesture.

Statistical Modeling

We used R software (R team, 2023), the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) for the generalized linear mixed model (GLMM), lmerTest for tests in linear mixed effects models (Kuznetsova, Brockhoff, & Christensen, 2017), and MuMIn for the model selection (Bartoń, 2020).

Firstly, to examine the combination between speech and gesture, we investigated whether participants' path gesture such as direction and trajectory expressed Motion phase among conditions such as context (beginning / ending) and Structure type (fall-cat and cat-photo structure) that included Verb information (Fall, Fly). In addition, each condition was coded using dummy coding then centered such as effect coding (e.g., −0.5, 0.5).

We explored the relationship between Verb expression and Motion phase by investigating the use of a gesture path, including motion trajectory and direction, utilizing generalized linear mixed models (GLMMs). Gesture occurrences were coded for motion trajectory (arc or straight) and direction (vertical or horizontal) using a frame-by-frame method. As for the model prediction of gesture trajectory, a maximum model was constructed, incorporating experimental conditions and their interactions as fixed effects, and individual and item differences as random effects. The model selection, using a forward stepwise approach, suggested that Structure type, Verb, Motion phase, and their interactions should be applied as fixed effects (glmer (*Trajectory (Arc)* ~ Verb * (Structure type + Motion phase) + Structure type: Motion phase + (1|Participants) + (1|Item))). Regarding the model prediction of gesture direction, it did not converge due to the correspondence between "vertical" and "horizontal" directions and the verb expression (vertical gestures: 70 cases within 70 "fall" verbs; horizontal gestures: 64 cases within 64 "fly" verbs). Consequently, the GLMM was not applied to data related to gesture direction. Instead, in the analysis of gesture direction, we examined differences between Structure type and Motion phases using a chi-square test.

To examine gesture onsets in a time series, we analyzed gesture production, especially path gestures, using cluster-based permutation analysis (CPA). This method, commonly applied in studies examining brain activation with techniques such as electroencephalography (EEG) and near-infrared spectroscopy (NIRS), was utilized to assess the time series of path gestures. First, the time course of gesture data, including conditions, was computed by binning at 100-millisecond intervals for the CPA, which necessitates high-density data in the time series. Since the time length of utterance differed among participants, we binned gesture data based on when the participant uttered the "first noun." We then specified that the target data be analyzed to compare with the temporal information conditions. Additionally, temporal information conditions were coded using effect coding (Ending = 1, Beginning = −1). We computed the CPA via the GLMM (Generalized Linear Mixed Model) using the "*clusterperm.glmer*" function with a binomial distribution (Voeten, 2018). This GLMM applied the Motion phase condition as a fixed effect, and individual and item differences as random factors.

Results

Motion Trajectory

The GLMM fit between the occurrence of gesture trajectory when the participants uttered the verb expression and each condition revealed that the Structure type (β = 4.013, z = 4.722, p < .001), the Verb (β = -1.549, z = 2.050, p = .040), the Motion phase (β = 1.468, z =2.134, p = .033), and the interaction of Structure type × Motion phase (β = -2.925, z = -2.031, p = .042) and Verb × Motion phase (β = 3.983, z = 2.434, p = .015) were significant (Figure 3).

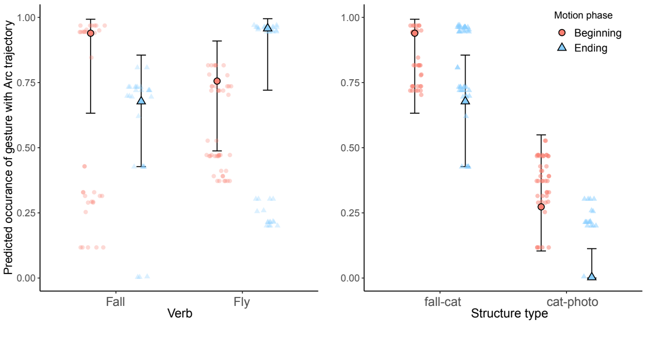

Figure 3: Predicted arc gesture trajectory for each condition.

To test the simple effects, simple contrasts were computed using dummy coding for each factor. The results revealed that in the picture associating to the cat-photo structure, arc trajectory gestures occurred more frequently at the beginning of Motion phase than at the ending (β = 2.931, z = 2.587, p

= .0097; Figure 3 Right). However, for the fall-cat structure, there was no significant difference in the occurrence of arc trajectory gestures between these motion phases. This result suggests that the falling of a cat photo was described at the beginning of the motion phase because the falling of a photo can be more readily described with arc gesture. Regarding the Verb "fall," arc trajectory gestures occurred more frequently at the beginning of Motion phase than at the end (β = 3.46, z = 2.567, p = .01). However, for the verb 'Fly," there was no significant difference in the occurrence of arc trajectory gestures between the beginning and the ending of Motion phase. This result suggests that at the beginning phase, arc gesture was used more when the verb "Fall" was uttered.

The Gesture Direction

Regarding gestures with vertical directions when participants uttered the "fall" Verb, the chi-square test revealed a significant association between Structure type and Motion phase (χ^2(1, *Case*= 64) = 8.347, p = .0038). However, in gestures with horizontal direction when participants uttered the "fly" Verb, the chi-square test tended to be a significant trend between Structure type and Motion phase (χ^2(1, *Case*= 70) = 3.673, p = .055).

In pictures containing the fall-cat structure, gestures with vertical directions occurred more frequently at the ending of Motion phase than at the beginning (*adj. residual* = ±3.16). Additionally, in pictures containing the cat-photo structure, gestures with vertical directions occurred more frequently at the beginning of Motion phase than at the ending (*adj. residual* = ±3.16). Gestures with horizontal directions, when participants uttered the "fall" Verb, tended to exhibit a similar trend to the results observed in gestures with a vertical direction.

Time Series of Gesture Trajectory

To compare gestural expressions, especially production of path gestures in their time courses, CPAs were computed using the *clusterperm.glmer* function in permutes. Figure 4 shows the time course of the results of the CPA based on GLMMs.

In the case of "arc" path gestures, participants depicted when they watched the picture containing left-branching information (i.e., fall-cat structure) with the verb "fall" prompted, specifically during the beginning motion phase of the picture (1400 to 2000 ms, cluster mass stat = 436.6). Additionally, when the verb "fly" was prompted, participants depicted using these gestures during both the beginning motion phase (-2300 to -1800 ms, cluster mass stat = 374.2) and the ending motion phase of the picture (1900 to 2800 ms, cluster mass stat = 623.7).

Moreover, when participants watched the picture containing right-branching information (i.e., cat-photo structure) with the verb "fall" prompted, they also depicted using the "arc" path gestures during the beginning motion phase (-1800 to 200 ms, cluster mass stat = 698.4). Additionally, participants also depicted using these gestures during the ending motion phase when the verb "fly" was prompted (-1700 to -500 ms, cluster mass stat = 410.3).

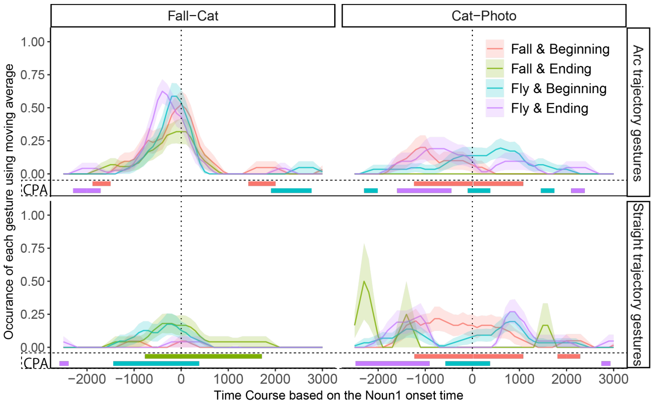

Figure 4: Time course of the gesture trajectory in each condition based on CPA using GLMM. Each color bar represents a significant cluster, indicating a significant occurrence compared to the motion-phase condition. Additionally, each color corresponds to a specific condition.

In the case of "straight" path gestures, participants depicted when they watched the picture containing left-branching information (i.e., fall-cat structure) with the verb "fall" prompted, specifically during the ending motion phase of the picture (-800 to 1800 ms, cluster mass statistic = 1535.2, p < .05). Additionally, when the verb "fly" was prompted, participants depicted using these gestures during both the ending motion phase (-3000 to -2500 ms, cluster mass statistic = 374.2, p < .05) and the beginning motion phase of the picture (-1500 to 400 ms, cluster mass statistic = 825.3, p < .05). Moreover, when participants watched the picture containing right-branching information (i.e., cat-photo structure) with the verb "fall" prompted, they also depicted using the "arc" path gestures during the beginning motion phase (-1300 to 1100 ms, cluster mass statistic = 831.4, p < .05). Additionally, when the verb "fly" was prompted, participants also depicted using these gestures during both the ending motion phase (-1700 to -500 ms, cluster mass stat = 410.3) and the beginning motion phase (-600 to 400 ms, cluster mass statistic = 762.3, p < .05).

These results indicate that, in the cat-photo structure, both arc path gestures and straight path gestures resemble the timings of occurrences. In contrast, in the fall-cat structure, the timings of gestures differed between arc path gestures and straight path gestures.

Discussion

The first aim of this study was to examine whether people produce co-speech gestures that express the meaning of verbs in accordance with different motion phases. We used two different phases of a motion event, beginning and ending. The second aim of this study was to examine whether people properly express relevant agents of the verb according to different situations. As for the second aim, we utilized ambiguous hierarchical structures that can be interpreted with two different agents, using two verbs that were expected to be described with different paths of gestures. The paths of gestures were analyzed according to the direction (vertical or horizontal) or trajectory (arc or straight).

Based on the analysis of spontaneous co-speech gestures, our first research question was answered: People do use co-speech gestures that express the meaning of verbs in accordance with different motion phases. The second research question was also answered: People do use different iconic gestures depending on the agents they wanted to describe. Information on motion phases and agents were simultaneously described in movements and hand shape of iconic gestures. People used more "arc" gestures when they described the beginning motion of the agent in the fall-cat structure. This result suggests that when people depicted the cat's falling, they tended to use arc gesture more often than when they depicted the cat-photo's falling. In addition, people tended to use more arc gesture to describe the beginning of a motion phase than the ending of it. It seems to suggest that people tried to depict the falling motion itself at the beginning of the motion, but they did not depict the falling motion when the motion was close to the end.

Thus, these results showed that the participants' use of co-speech gestures differed when they expressed different motion phases of beginning or ending, and associated agents were also described with different paths of gestures. In addition, our permutation analysis focusing on two different types of verbs further supported the nature of co-speech gestures regarding verbs. While the speech prompt was kept consistent, co-speech gestures showed variations, suggesting that gestures can compensate the motion phases and associated agents that may not be described by language.

Acknowledgments

We would like to thank all the participants in this study. This study was supported by JSPS/MEXT KAKENHI JP17H06382 (H.K.), JSPS KAKENHI JP20H01763 (H.K.), and JP20K03375 (T.Y.).

References

Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of manner and path: A comparison of English, Japanese, and Turkish. *Cognition*, 102(1), 16-48.

Bartoń, K. (2022). *MuMIn*: Multi-Model Inference. R package version 1.47.5.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637 /jss.v067.i01

Bull, P., & Connelly, G. (1985). Body movement and emphasis in speech. *Journal of Nonverbal Behavior*, 9(3), 169–187.

Duncan, S. D. (2002). Gesture, verb aspect, and the nature of iconic imagery in natural discourse. Gesture, 2(2), 183-206.

ELAN (Version 6.7) [Computer software]. (2023). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, 5(6), 664–674.

Hirsh-Pasek, K., & Golinkoff, R. M. (Eds.). (2010). Action meets word: *How children learn verbs*. Oxford University Press.

Handa, Y., Yasuda, T., & Kobayashi, H. (2021). The use of co-speech gestures in conveying Japanese phrases with verbs. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43, 1555–1559.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16(5), 367–371.

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., 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.

Kendon, A. (2004). Gesture: Visible action as utterance. Cambridge University.

Kimura, H., Yasuda, T., & Kobayashi, H. (2023). Spontaneous co-speech gestures with prompt phrases reflect linguistic structures. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

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, 16–32.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). *lmerTest Package*: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13), 1-26.

Lüdecke, D. (2018). *ggeffects*: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, 3(26), 772. https://doi.org/10.21105/joss.00772

R Core Team (2023). *R*: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Shirai, Y. (2000). The semantics of the Japanese imperfective-teiru: An integrative approach. Journal of pragmatics, 32(3), 327-361.

Voeten, C. C. (2018). *permutes*: Permutation tests for time series data. R package version 0.1. Available online at: https://CRAN.R-project.org/package=permutes

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D.A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., ... & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.