An experimental investigation of perspective alignment in gesture and speech

Summary. Hinterwimmer et al. (2021) experimentally investigated the hypothesis that perspective in gesture and speech are by default aligned, i.e., when a character's or protagonist's perspective is conveyed in the speech signal, this utterance is preferably aligned with a *character viewpoint gesture* (CVG). If an utterance expresses an observer's perspective, by contrast, it is more likely accompanied by an observer viewpoint gesture (OVG). Their results, however, showed an overall preference for CVGs. They argued that there were pragmatic factors (e.g., informativity) at play blocking the hypothesized perspective alignment. The study reported here further investigates Hinterwimmer et al.'s (2021) hypothesis by comparing two different CVGs paired with a verbal utterance in a rating study. The results suggest that, contrary to Hinterwimmer et al.'s (2021) hypothesis, multiple perspectives can be simultaneously expressed in gesture and speech.

Background. Normally, an utterance expresses the speaker's perspective or viewpoint. Therefore, all perspective-dependent expressions (e.g., relational expressions such as *left* and *right*) are by default interpreted from the speaker's perspective (e.g., Harris and Potts, 2009). It is possible, however, to shift the perspective from the speaker to some other individual which is salient in the current discourse. Examples are instances of *reported speech*. Perspective can also be expressed in gesture (McNeill, 1992). A common distinction in this line of research is the one between CVGs and OVGs. CVGs depict an event from a first-person perspective, OVGs depict an event from a third-person perspective. There is very little research on how perspective taking in the two modalities interacts. Hinterwimmer et al. (2021) posited the hypothesis that the perspectives expressed in gesture and speech should be aligned. They ground their hypothesis on previous research which has found that i) gesture and speech convey a joint multimodal message which is planned by one central cognitive process and later passed on to different communication channels (e.g., de Ruiter, 1998) and ii) perspective in gesture and speech have the same conceptual source (Parrill, 2010). In order to test their hypothesis, they designed a forced-choice study where they paired videotaped utterances in *free indirect discourse* (FID), which clearly expressed a salient protagonist's perspective, or a more general statement describing an event from an observer's perspective with a CVG and an OVG. Participants then had to select the version of the utterance which they considered more natural. They predicted that CVGs were preferred in the FID condition, while OVGs were predicted to be preferred in the condition where the event was described from an observer's perspective. However, contrary to their hypothesis, they found an overall preference for CVGs regardless of the perspective expressed in speech. Hinterwimmer et al. (2021) hypothesized that this might be due to pragmatic factors which block the default perspective alignment, e.g., that CVGs are more salient than OVGs due to their differences in size. This was experimentally validated by Walter et al. (2023). Therefore, the hypothesis that perspective in gesture and speech are preferably aligned is difficult to investigate when comparing CVGs and OVGs. It thus seems more promising to investigate cases where one can compare two occurrences of the same type of viewpoint gesture. In sentences where two protagonist's perspectives are introduced, it is therefore hypothesized that a CVG conveying the more prominent protagonist's perspective is preferred over a CVG conveying the less prominent protagonist's perspective.

Experimental study. An experimental rating study was conducted (2x2 design) in order to investigate this hypothesis. In each item, two protagonist's perspectives were introduced: a prominent one and one which was less prominent. The sentences were either aligned with a CVG from the more prominent protagonist's perspective or a CVG conveying the less prominent protagonist's perspective (factor Gesture). Moreover, the more prominent perspective was either introduced from a first-person perspective via a first-person pronoun or from a third-person perspective via a proper name (factor Referential Expression). The second protagonist was always introduced by an indefinite. 24 experimental items were construed along the lines of the example in (1). The exper-

imental items were split up according to a Latin square design and interspersed with 25 fillers. 40 native speakers of German were recruited via Prolific for participation. Participants had to rate the items on a 7-point Likert scale for naturalness (1 = completely unnatural; 7 = completely natural).

(1) a. Gestern Abend ist mir etwas Krasses passiert. Ich war im Park spazieren und auf einmal kam ein Typ auf mich zu und hat mich ohne Vorwarnung so heftig geschubst, dass ich fast hingefallen wäre, weil ich das Gleichgewicht verloren habe.

b. Gestern Abend ist Paula etwas Krasses passiert. Sie war im Park spazieren und auf einmal kam ein Typ auf sie zu und hat sie ohne Vorwarnung so heftig geschubst, dass sie fast hingefallen wäre, weil sie das Gleichgewicht verloren hat.

'Yesterday evening something crazy happened to me/Paula. I/she was taking a walk in the park when suddenly some guy walked to me/her and nudged me/her so strongly that I/she nearly fell because I/she lost my/her balance.'

Prominent CVG: Speaker is staggering backwards and flailing about.

Not prominent CVG: Speaker performs a nudging gesture.

In (1) the CVG conveying the backwards staggering aligns with the perspective which is more prominent on the level of the speech signal, since it conveys the speaker's (1a) or Paula's perspective (1b) on the described event. It should therefore be preferred over the nudging CVG, which expresses the perspective which is less prominent on the level of the speech signal. Based on the hypothesis that there is perspective alignment in gesture and speech, a main effect for Gesture is predicted. Moreover, since introducing a perspective by means of a first-person pronoun makes that perspective even more prominent, an interaction between the two factors is predicted. The results show that the conditions were all rated equally well (first-person + prominent CVG: M = 5.43, SD = 1.53; first-person + not prominent CVG: M = 5.39, SD = 1.47; proper name + prominent CVG: M = 5.47, SD = 1.43; proper name + not prominent CVG: M = 5.33, SD = 1.53). An ordinal mixed-effects model was fitted onto the data and yielded neither a main effect for the factor Gesture nor significant interactions.

Discussion and conclusion. The results show that both CVGs were equally acceptable regardless of the prominent perspective in the speech signal. Moreover, the factor Referential Expression did not have any influence on the ratings either. The results thus do not confirm the hypothesis that perspective in gesture and speech are preferably aligned. In contrast to Hinterwimmer et al.'s (2021) study there were no intervening pragmatic factors which might have blocked perspective alignment in this study. The most plausible conclusion is therefore to reject the hypothesis that there is a preference for perspective alignment in the two modalities. Rather, multiple perspectives can be simultaneously expressed. Future research should investigate whether there are any constraints for expressing multiple perspectives in the two modalities.

References

de Ruiter, J. P. (1998). *Gesture and Speech Production*. University of Nijmegen PhD Thesis.

Harris, J. A. and C. Potts (2009). Perspective-shifting with appositives and expressives. *L&P, 32(6)*.

Hinterwimmer, S., U. Patil, and C. Ebert (2021). On the interaction of gestural and linguistic perspective taking. *FiC, 6*.

McNeill, D. (1992). *Hand and Mind: What Gestures reveal about Thought*. UChicago Press.

Parrill, F. (2010). Viewpoint in speech-gesture integration: Linguistic structure, discourse structure, and event structure. *LaCP, 25, 5*.

Walter, S., C. Ebert, and S. Hinterwimmer (2023). Are there salience differences between character and observer viewpoint gestures? Poster at *XPrag X*.