An experimental investigation of perspective alignment in gesture and speech

Sebastian Walter & Stefan Hinterwimmer*

Abstract. Hinterwimmer et al. (2021) experimentally investigated the hypothesis that perspective in gesture and speech is 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*. If an utterance expresses an observer's perspective, by contrast, it is more likely accompanied by an *observer viewpoint gesture*. Their results, however, showed an overall preference for character viewpoint gestures. 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 character viewpoint gestures paired with a verbal utterance in a rating study. The results suggest that, contrary to Hinterwimmer et al.'s (2021) hypothesis, multiple, potentially non-aligned perspectives can be simultaneously expressed in gesture and speech.

Keywords. Perspective in gesture; free indirect discourse; interactions of perspective taking in gesture and speech

1. Introduction. Perspective plays a decisive role in the interpretation of many lexical items. The evaluative expression *fantastic* in (1), for example, belongs to the family of so-called *perspective-dependent expressions*.

(1) The conference was fantastic!

It is apparent that an evaluative expression as in (1) is by default interpreted from the speaker's perspective (Harris 2012). However, sometimes perspective-dependent expressions can occur in contexts where they do not depend on the perspective of the speaker of the current utterance situation:

(2) a. Marvin said: "The conference was fantastic!"
b. Marvin said that the conference was fantastic.

In the *direct discourse* utterance (2-a) as well as the *indirect discourse* utterance (2-b), *fantastic* is interpreted from Marvin's rather than the speaker's perspective.

Perspective can also be expressed in gesture (McNeill 1992). The main distinction to be drawn here is the one between *character* and *observer viewpoint gestures*. Character viewpoint gestures depict an event from an internal, i.e., first-person perspective. Observer viewpoint gestures, by contrast, depict events from a more external, third-person perspective. Imagine that a speaker

*Acknowledgments: This research was conducted in the DFG-funded project *Visual and non-visual means of perspective taking in language* which is part of the Priority Program 2392 (*ViCom*). We thankfully acknowledge the financial support of the German Research Foundation (DFG). Moreover, we would like to thank our actor Lennart Klappstein for the enactment of the experimental material of the study reported in this paper. Moreover, we would like to thank our student assistant Lennart Fritzsche for helping with the video recordings. Finally, we express our special gratefulness to Cornelia Ebert for discussing the topics covered in this paper. Authors: Sebastian Walter, University of Frankfurt (s.walter@em.uni-frankfurt.de) & Stefan Hinterwimmer, University of Hamburg (stefan.hinterwimmer@uni-hamburg.de).

wants to describe an event where someone had to run. When they want to include a gesture in that description, they could either perform a character viewpoint gesture where they depict the running with their whole body (i.e., as if they were running on one spot). Alternatively, they could produce an observer viewpoint gesture where they just use their index and middle finger to represent the running person's legs and thus depict the running event from an external perspective.

Research on interactions of perspective taking in gesture and speech is scarce. Hinterwimmer et al. (2021) report a study investigating the hypothesis that perspective in gesture and speech is by default aligned. Contrary to their hypothesis, however, the results showed an overall preference for character viewpoint gestures. They conclude that there might have been pragmatic factors blocking perspective alignment in their experimental stimuli. The study reported in this paper controls for these intervening pragmatic factors. Contrary to the hypothesis, however, the results still suggest no strict preference for perspective alignment in gesture and speech. Instead, multiple, potentially non-aligned perspectives can be expressed simultaneously in the two modalities.

The paper is structured as follows: Section 2 will provide the relevant background on perspective in speech, gesture, and previous research on interactions of the expression of perspective in the two modalities. In Section 3, the study is described which investigates Hinterwimmer et al.'s (2021) weakened hypothesis that in the absence of intervening pragmatic factors, perspective in gesture and speech is aligned. Finally, Section 4 offers a general discussion of the results and gives implications for future research.

2. Background.

2.1. PERSPECTIVE IN SPEECH. The perspective conveyed in an utterance is normally that of the speaker. Therefore, perspective-dependent expressions, such as *epithets* (e.g., *that bastard*), *relational expressions* (e.g., *left*, *right*, *this*, *that*), *predicates of personal taste* (e.g., *tasty*), or *epistemic modals* (e.g., *must*) are by default interpreted from the speaker's perspective (Harris & Potts 2009, Harris 2012). However, one can find systematic exceptions to this general tendency, namely instances of *reported speech*:

(3) a. On her way home, Mary heard a song by Kendrick Lamar that she liked on the radio. She thought: "I will buy his new album tomorrow."
b. On her way home, Mary heard a song by Kendrick Lamar that she liked on the radio. She thought that she would buy his new album on the following day.
c. On her way home, Mary heard a song by Kendrick Lamar that she liked on the radio. She would buy his new album tomorrow.

(Hinterwimmer 2017, p. 284)

In the instance of direct discourse (DD) in (3-a), all perspective-dependent expressions shift toward the reported speaker and are thus interpreted from their perspective, i.e., Mary's. This means, then, that the first-person pronoun *I* refers to Mary, not the speaker in the current utterance situation. By contrast, the picture for indirect discourse (ID) and *free indirect discourse*, FID, ((3-b) and (3-c), respectively) is somewhat less clear.

In ID (cf. (3-b)) epithets, evaluative expressions, and predicates of personal taste can in principle either be evaluated from the speaker's perspective or from the perspective of the matrix clause subject (Mary), the latter being the default option. Personal pronouns (in many languages, but see

Anand & Nevins 2004), temporal, and local deictic expressions are also evaluated from the matrix subject's perspective. For the former two, however, this preference can be easily overwritten (Plank 1986, Anderson 2019). By contrast, expressives and appositives are normally interpreted from the speaker's perspective, but a shifted interpretation is also possible (Harris & Potts 2009).

Finally, the second sentence in (3-c) is an instance of FID, a way to report a protagonist's thoughts or speech without any overt marking (e.g., Hinterwimmer 2024). Here, all perspective-dependent expressions are interpreted from the protagonist's perspective with only two exceptions: pronouns and tenses (e.g., Schlenker 2004). This explains why the past tense marking can co-occur with the temporal adverbial *tomorrow* in (3-c) without resulting in ungrammaticality or at least a contradictory interpretation. Therefore, the protagonist's and the speaker's perspective are conveyed in FID although the protagonist's perspective is more prominent.

2.2. PERSPECTIVE IN GESTURE. Beside encoding perspective in spoken and written language, *co-speech gestures* have also been shown to encode perspective (McNeill 1992, among others). Verbal utterances are often accompanied by gestures, which can either be manual, i.e., performed with the hands and potentially other body parts, or facial, i.e., performed with the face. These gestures are often synchronized with the verbal expressions they co-occur with. The stroke (= the core) of a gesture, for instance, is usually aligned with the nuclear accent of a word (Loehr 2004, Ebert et al. 2011). Moreover, different alignment patterns of gesture and speech have been shown to have different semantic effects (Ebert & Ebert 2014). This claim has been experimentally validated by the study reported in Ebert et al. (2022).

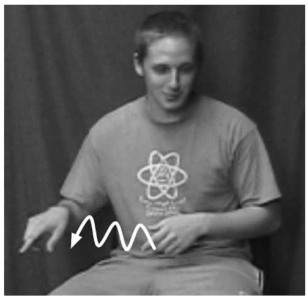
Previous research has distinguished different gesture types (for an overview, see McNeill 1992), among them *iconic gestures*. Iconic gestures visually resemble a property of an object or action they illustrate. Perspective is often encoded in iconic gestures. McNeill (1992) distinguishes between character viewpoint gestures (CVGs) and observer viewpoint gestures (OVGs, see also Parrill 2010, Stec 2012, among others).

CVGs illustrate an event from a first-person perspective and the whole body is usually involved in the production of the gesture. OVGs, by contrast, illustrate an event as if observed from a distance (i.e., from a third-person perspective) and therefore only the hands are involved when producing the gesture. CVGs have been argued to be more informative than OVGs (Beattie & Shovelton 2002). Examples are given in Figures 2a and 2b, respectively. Both are taken from the study reported in Parrill (2010) where participants had to describe cartoon scenes as in Figure 1 to friends who had not seen the clips. When describing the hopping movement of the skunk as shown in Figure 1, speakers have several options to also incorporate gesture. The gesture shown in Figure 2a is a clear instance of a CVG as the speaker depicts the skunk's movement from a first-person and thus an internal perspective by imitating the skunk's posture and also to a certain extent its facial expression. On the other side, the speaker shown in Figure 2b produces a fairly clear instance of an OVG as they only trace the skunk's trajectory and hopping with their index finger, thus adopting a more external, third-person perspective.

There is a third type of viewpoint gesture, which occurs very infrequently, however. This gesture type encodes multiple viewpoints at the same time and has therefore been dubbed *dual viewpoint gesture* (Parrill 2009). Encoding multiple viewpoints in gesture seems to be less constrained than the occurrence of multiple viewpoints in speech since dual viewpoint gestures allow

Figure 1: Cartoon scene of a skunk hopping across a room. Taken from Parrill (2010)

(a) CVG used to depict the skunk in Figure 1. (Figure 3 in Parrill 2010, p. 652)



(b) OVG used to depict the skunk in Figure 1. (Figure 2 in Parrill 2010, p. 651)

Figure 2: Examples of a CVG and an OVG to depict the event shown in Figure 1

for the presence of two (equally prominent) character viewpoints at the same time, which has not been attested for spoken or written language. Interestingly, this is also possible in sign languages (e.g., Maier & Steinbach 2022). Therefore, this might be a modality-specific feature. The co-presence of a character's and an observer's viewpoint in a gesture has been argued to be possible although it often produces an ironic effect because the two viewpoints seem to compete with each other (McNeill 1992). Moreover, dual CVGs are restricted to specific contexts at least for adults (McNeill 1992). More specifically, one of the CVGs always is a deictic gesture to the speaker's body, representing the viewpoint of one character, and the body represents another character viewpoint. It can be noted, in sum, that the expression of multiple viewpoints is more liberal in gesture as opposed to speech.

2.3. INTERACTIONS OF PERSPECTIVE-TAKING IN GESTURE AND SPEECH. Expressing viewpoint in gesture is not entirely independent of the verbal material the gesture co-occurs with. Parrill (2010), for example, has noted that an interdependence of linguistic, event, and discourse structure affects the choice of the gestural viewpoint. New information has been argued to co-occur with CVGs more frequently than with OVGs (McNeill 1992, Parrill 2010). Transitive utterances are more frequently accompanied by a CVG than by an OVG. Moreover, events in which the speaker shows affect are also more likely to be accompanied by a CVG. It has furthermore been noted that events in which a trajectory is described are more likely to be accompanied by an OVG. Crucially, however, viewpoint in gesture and speech have been argued to share a conceptual source (Parrill

2009, 2010).

In a similar vein, Kita & Özyürek (2003) present findings from crosslinguistic work suggesting that the execution and also the form of gestures depend on a language's lexical and structural encoding. In other words, gesture execution and gesture form depend on the language of the speaker. If, for example, trajectory is encoded in a word (e.g., the English noun *swing*), a gesture co-occurring with this word is also more likely to encode trajectory. This constitutes a clear interaction of linguistic form and gesture execution, thus suggesting that there also is an interaction between linguistic and gestural viewpoint.

The study reported in Hinterwimmer et al. (2021) investigates the hypothesis that perspective in gesture and speech is by default aligned. The authors develop this hypothesis based on the assumption that gesture and speech convey a multimodal message which is planned by one central cognitive process. This message is then passed on to different communication channels (cf. McNeill 1992, de Ruiter 1998, among others). Moreover, as has been laid out above, viewpoint expressed in gesture and speech has the same conceptual source (Parrill 2010) and perspective is an essential part of multimodal messages. Finally, they claimed for the information conveyed in the two communication channels to be coherent by default (although gesture-speech mismatches can sometimes be useful, cf. Goldin-Meadow 1999), thus allowing them to straightforwardly derive the above-stated hypothesis.

In order to test for their hypothesis, they constructed stimuli which either expressed the perspective of an individual participating in the event, i.e., a protagonist's perspective, or the perspective of the speaker, i.e., a narrator's perspective. These utterances were then paired with CVGs and OVGs. An example can be found in (4). Underlined parts of an example indicate gesture-speech alignment.

(4) a. **Narrator's perspective:** Leon ist ein begeisterter Sportler. Als er sich neulich beim Fußballspielen den Ball erkämpfte, kickte er ihn sofort in Richtung Tor. + CVG/OVG
'Leon is an enthusiastic athlete. When he recently won the ball while playing soccer, he immediately kicked it in the direction of the goal.'

b. **Character's perspective:** Leon spielte am Wochenende Fußball. Nach einigem Gerangel hatte er sich den Ball erkämpft. Toll, jetzt konnte er ihn direkt in Richtung Tor schießen! + CVG/OVG
'Leon played soccer on the weekend. After some scramble, he had finally won the ball. Great, now he could directly kick it in the direction of the goal!'
CVG: Speaker performs a kicking movement with their right leg and foot with an enthusiastic facial expression.
OVG: Speaker presses their index finger on their thumb, then releasing it quickly, thus imitating a kicking movement with their index finger.

(Hinterwimmer et al. 2021, p. 8)

Participants were presented either both versions of a stimulus in the narrator's perspective, meaning that they saw videotaped versions of the same verbal utterance, once accompanied by a CVG and once accompanied by an OVG, or they saw both versions of the verbal stimulus in FID, i.e., in the protagonist's perspective. Hinterwimmer et al. (2021) used a forced-choice paradigm for the

study, meaning that participants had to choose the preferred version of the stimulus from the two available options. They predicted that stimuli expressing a protagonist's perspective on the speech level should be preferred when they are accompanied by a CVG. When a stimulus expressed a narrator's perspective, by contrast, the OVG was predicted to be preferred.

Contrary to their predictions, the results showed an overall preference for the CVG items regardless of the perspective expressed in the speech signal. There are two potential explanations for Hinterwimmer et al.'s (2021) findings: i) there is no preference for perspective alignment in multimodal messages or ii) the default for linguistic and gestural perspective indeed is to be aligned, but intervening pragmatic factors can overwrite this preference. Hinterwimmer et al. (2021) note that most of their stimuli contained transitive utterances as target sentences. As has been pointed out above, previous research has shown that CVGs tend to be preferred over OVGs in transitive utterances and that CVGs are more likely to be used when new information is expressed verbally (McNeill 1992, Parrill 2010), which could thus have caused the overall preference for CVGs. Moreover, CVGs are in general more informative than OVGs (Beattie & Shovelton 2002), which could be a further reason why CVGs were preferred over OVGs. The two types of viewpoint gestures differ in terms of size. While the whole body is usually involved in the production of a CVG, only hands and arms are involved in the production of an OVG (McNeill 1992). This size difference potentially makes CVGs more salient than OVGs (but see Walter 2024 for tentative evidence against this claim), which could in turn also account for the observed CVG preference.

In order to control for the potential pragmatic factors which might have blocked perspective alignment, the follow-up rating study reported in Section 3 was conducted where only CVGs were used as gestures. Using only CVGs controls for the aforementioned factors because they do not differ in terms of size and informativity, for example. This time, items were constructed where two protagonist's perspectives were introduced on the speech level. One of them was prominent, the other one was not. In one condition, a CVG co-occurred with the speech signal that matched the prominent protagonist's perspective on the speech level. In the other condition, a CVG was aligned with the verbal stimulus matching the non-prominent protagonist's perspective. Following Hinterwimmer et al. (2021), it was hypothesized that this time, the condition with the CVG matching the prominent protagonist's perspective on the speech level should be preferred as the factors potentially blocking perspective alignment discussed above were controlled for.

3. Experimental study.

3.1. Method.

3.1.1. Participants. Self-reported native speakers of German (n = 40) were recruited via Prolific. They were naive with respect to the research question.

3.1.2. Materials. For materials, 24 experimental items were constructed which were then videotaped. Each item consisted of two sentences: the first one introduced an event and the second one further elaborated on that event. In the first sentence, a protagonist was introduced whose perspective was made prominent on the speech level and picked up again in the second sentence. A further, non-prominent protagonist's perspective was introduced in the second sentence of each experimental item. In order to keep this perspective less prominent, this referent was always introduced by means of an indefinite DP (see Meuser et al. to appear and Meuser 2022 for experimental

evidence that referents introduced by indefinite DPs are less prominent as perspective-takers than referents introduced by proper names). Crucially, the prominent protagonist's perspective was either introduced by means of a first-person pronoun or a proper name (factor REFERENTIAL EXPRESSION). The utterances were accompanied either by a CVG matching the prominent protagonist's perspective or by a CVG matching the non-prominent protagonist's perspective (factor GESTURE). The study was thus of a 2x2 design. An example item is given in (5).

(5) 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$_{\text{not prominent}}$, dass ich fast hingefallen wäre$_{\text{prominent}}$, weil ich das Gleichgewicht verloren habe.
'Yesterday evening something crazy happened to me. I was taking a walk in the park when suddenly some guy walked to me and nudged$_{\text{not prominent}}$ me so strongly that I nearly fell$_{\text{prominent}}$ because I lost my balance.'

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$_{\text{not prominent}}$, dass sie fast hingefallen wäre$_{\text{prominent}}$, weil sie das Gleichgewicht verloren hat.
'Yesterday evening something crazy happened to Paula. She was taking a walk in the park when suddenly some guy walked to her and nudged$_{\text{not prominent}}$ her so strongly that she nearly fell$_{\text{prominent}}$ because she lost her balance.'
Prominent CVG: Speaker is staggering backwards and flailing about.
Not prominent CVG: Speaker performs a nudging gesture.

In order to distract participants from the research question, the experimental items were interspersed with 25 unrelated fillers. In addition, there was a training session consisting of two items.

3.1.3. PROCEDURE. Before the training session, participants were made familiar with the task by an introductory text. In this text, they were also instructed to pay attention to the audio as well as the videotape. Moreover, they were informed about their data protection rights and gave informed consent. The questionnaire was created using SoSci Survey (Leiner 2022), an online platform for creating questionnaires which can be used free of charge for academic purposes. The questionnaire was distributed via Prolific using its pre-filtering functions to exclusively select for German native speakers as participants. The lists of the questionnaire were run independently in order to be able to filter out those participants who had completed a previous list of the questionnaire. This was done to prevent participants from participating multiple times. The items were split up according to a Latin square design and evenly distributed onto four lists. The 25 fillers as well as the two training items were included on each list. Experimental items and fillers occurred in a randomized order for each participant. In order to check the participants' attention, they were asked to answer three questions about the videos after the final trial (e.g., they were asked what the hair color of the person they saw in the videos was). Participants had to rate on a 7-point Likert scale how natural they considered each utterance (1 = completely unnatural; 7 = completely natural).

3.2. PREDICTIONS. Based on the hypothesis that perspective in gesture and speech is aligned if there are no intervening pragmatic factors, the CVG representing the prominent protagonist's perspective should always be preferred over the CVG representing the non-prominent protagonist's perspective. Moreover, since introducing a perspective by means of a first-person pronoun

makes this perspective more prominent compared to introducing it by means of a proper name (see Bimpikou 2020 and Saure et al. 2023 for experimental evidence for this assumption), the preference for the CVG matching the prominent protagonist's perspective should be higher in the former case compared to the latter. Thus, an interaction between the factors REFERENTIAL EXPRESSION and GESTURE is predicted.

3.3. RESULTS. The data was analyzed using the R statistics software (R Core Team 2022). To test for significant effects, the results were analyzed using a cumulative link mixed effects model with the clmm() function in the R package ordinal (Christensen 2023). For the analysis, an ordinal mixed effects model was chosen instead of a linear mixed effects model out of two reasons: first, linear mixed effects models require normally distributed data and second, they require the data to be measured at the interval level. Both is questionable for Likert scale data, making an ordinal mixed effects model the more adequate choice. The two factors were entered into the model as fixed effects using effect coding, that is the intercept represents the unweighted grand mean and the fixed effects compare the factor levels to each other. The full analysis script as well as the materials can be found here: https://osf.io/7hwjz/.

The mean values and standard deviations (SDs) are shown in Figure 3. In general, there are only very subtle rating differences in the ratings for the CVG matching the prominent protagonist's perspective (first-person pronoun: $M = 5.43$, $SD = 1.53$; proper name: $M = 5.47$, $SD = 1.43$) as well as for the CVG matching the non-prominent protagonist's perspective (first-person pronoun: $M = 5.39$, $SD = 1.47$; proper name: $M = 5.33$, $SD = 1.53$). The results also show that there were no big rating differences in general between the CVG matching the prominent protagonist's perspective and the CVG matching the non-prominent protagonist's perspective. The ordinal mixed effects model corresponding to the data shown in Figure 3 is given in Table 1. Unsurprisingly, neither main effects nor interactions are observable in the model output.

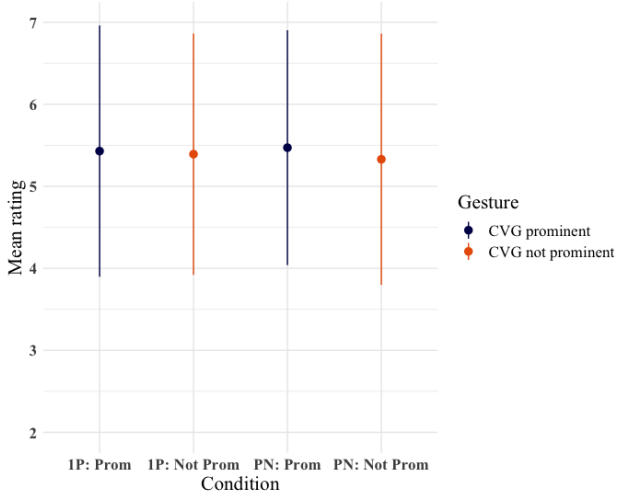


Figure 3: Mean values and standard deviations for each condition. (Abbreviations: 1P = First-person pronoun, PN = Proper name, Prom = CVG from the prominent protagonist's perspective, Not Prom = CVG from the not prominent protagonist's perspective)

	Estimate	Std. Error	z value	Pr(> |z|)
Gesture	.177	.121	1.472	.141
Referential expression	-.075	.121	-.619	.536
Gesture:Refrential expression	.064	.241	.264	.792

Table 1: Ordinal mixed-effects model with Mode and Match as fixed effects and participants and items as random intercepts. Formula: choiceO ∼ Gesture * RefExp + (1|CASE) + (1|item) Significance codes: *** 0.001 | ** 0.01 | * 0.05 | . 0.1

3.4. DISCUSSION. In contrast to Hinterwimmer et al.'s (2021) seminal study investigating alignment patterns of perspective in gesture in speech, the study reported in this paper controlled for intervening pragmatic factors which potentially block perspective alignment since in this study only CVGs were compared. However, the model output still does not confirm the hypothesis that perspective in gesture and speech is aligned in the absence of intervening pragmatic factors because this hypothesis straightforwardly translates to predicting an interaction between the factors REFERENTIAL EXPRESSION and GESTURE. This interaction cannot be found in the data. Furthermore, not even a main effect of GESTURE can be observed, indicating that the gesture matching the non-prominent protagonist's perspective was equally preferred as the the gesture matching the prominent protagonist's perspective.

4. General discussion and conclusion. The study reported in Hinterwimmer et al. (2021) (cf. Section 2.3 of this paper) for the first time investigated alignment patterns of gestures and speech. Based on i) findings that a multimodal message is planned by one underlying cognitive process (e.g., de Ruiter 1998), ii) findings that viewpoint in gesture and speech have the same conceptual source (Parrill 2010), and iii) the assumption that information conveyed in the two communication channels is by default coherent, the authors hypothesized that perspective in gesture and speech is by default aligned. They investigated this in a forced-choice study comparing CVGs and OVGs. The results, however, showed an overall CVG preference although this preference was slightly (but not significantly) smaller in the condition where the narrator's perspective was prominent on the speech level. Based on this slightly smaller preference, they argued that there might have been pragmatic factors that blocked the hypothesized perspective alignment. Examples for these pragmatic factors are the transitivity of the target sentences in their study (transitive events more likely evoke CVGs, cf. Parrill 2010) or salience differences between the two gesture types (but see Walter 2024 for results suggesting otherwise). They therefore concluded with the weakened hypothesis that perspective in gesture and speech is aligned, but only if there are no intervening pragmatic factors blocking the alignment.

The study reported in this paper investigated Hinterwimmer et al.'s (2021) weakened hypothesis by means of a rating study controlling for the aforementioned pragmatic factors which might have blocked the perspective alignment in their study. In this study, two protagonist's perspectives were introduced on the speech level, one of them being prominent and the other one being not prominent. These verbal stimuli were then either aligned with a CVG matching the prominent protagonist's perspective or a CVG matching the non-prominent protagonist's perspective. It was hypothesized that the CVG matching the prominent protagonist's perspective should be preferred.

This was not borne out by the results, however. Taking together the results reported in Hinterwimmer et al. (2021), both studies point toward rejecting the hypothesis as both times no alignment preference could be obtained from the data. Instead, it seems that multiple perspectives can be expressed relatively freely in gesture and speech. It remains an open question, however, if there are any constraints on which perspectives can be simultaneously expressed in the two modalities. Still, it might be a little premature to reject the perspective alignment hypothesis as will be laid out below in further detail.

The results of the studies are somewhat surprising, since there is empirical evidence in favor of a preferred perspective alignment from sign languages. In a study with signers of German and Turkish Sign Language, for example, it has been found that signers tend to adopt a character's perspective when producing a handling classifier, i.e., a classifier showing how an object was handled, and an observer's perspective for entity classifiers, that is classifiers denoting an entity from an outside perspective (Özyürek & Perniss 2011). Moreover, an overall preference for adopting a character's perspective for signers of German and Turkish Sign Language was observed in this study. This could in principle at least explain the CVG preference observed by Hinterwimmer et al. (2021). The preference to adopt a character perspective could then be a specific property of the visual modality also in spoken languages. A potential explanation for this character viewpoint preference is as follows: It has been noted at least for signers that they prefer depicting over describing when reporting an event (Engberg-Pedersen 1993, Özyürek & Perniss 2011). Since adopting a character viewpoint in gesture allows for richer depictions compared to adopting an observer viewpoint as the whole body is involved when adopting the former viewpoint (cf. McNeill 1992), the character viewpoint preference might come about because of a general preference for depicting over describing when reporting events. In general, the visual modality is more suitable than the acoustic modality to depict events. Assuming that the preference to depict when reporting events also holds for spoken languages, the observed preference for CVGs is easily explained. This, in turn, can also be transferred to the findings of the study reported in this paper: although there potentially is a preference for perspective alignment also in spoken languages, both CVGs used in the study were equally depictive. It is thus possible that the experimental items were not suitable to test for our hypothesis. In addition, the preference to adopt a character perspective in the visual modality and to therefore prefer CVGs in spoken language might be so strong that perspective alignment is only experimentally testable in cases where the perspective of a protagonist is made extremely prominent on the speech level. One of these cases are so-called *be like*-constructions where not only words, but also actions and other non-linguistic behavior can be quoted under a demonstrational account to quotation (Clark & Gerrig 1990, Davidson 2015).

(6) John was like hurrying to get to the bus station on time. + CVG depicting running

Be like-constructions thus have a highly depictive component while at the same time making a protagonist's perspective very prominent on the level of speech as they directly quote that protagonist's speech and/or actions. Intuitively, an alternation of (6) with an OVG depicting John's running instead of a CVG seems less natural than (6). This indicates that here, perspective alignment holds. Before ultimately rejecting the hypothesis of a preference for perspective alignment, this should be tested experimentally. We leave this to future research.

References

Anand, Pranav & Andrew Nevins. 2004. Shifty operators in changing contexts. In Robert B. Young (ed.), *Proceedings of Semantics and Linguistic Theory (SALT) 14*, 20–37. Ithaca, NY: CLC Publications. 10.3765/salt.v14i0.2913.

Anderson, Carolyn J. 2019. Tomorrow isn't always a day away. In M. Teresa Espinal, Elena Castroviejo, Manuel Leonetti, Louise McNally & Cristina Real-Puigdollers (eds.), *Proceedings of Sinn und Bedeutung 23*, 37–56. Barcelona, Spain: Universitat Autònoma de Barcelona.

Beattie, Geoffrey & Heather Shovelton. 2002. An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology* 93(2). 179–192. 10.1075/gest.1.2.03bea.

Bimpikou, Sofia. 2020. Who perceives? Who thinks? Anchoring free reports of perception and thought in narratives. *Open Library of Humanities* 6(2). https://doi.org/10.16995/olh.484.

Christensen, Rune H. B. 2023. *ordinal—Regression Models for Ordinal Data*. https://CRAN.R-project.org/package=ordinal. R package version 2023.12-4.

Clark, Herbert H. & Richard J. Gerrig. 1990. Quotations as demonstrations. *Language* 66(4). 764–805. 10.2307/414729.

Davidson, Kathryn. 2015. Quotation, demonstration, and iconicity. *Linguistics and Philosophy* 38(6). 477–520. 10.1007/s10988-015-9180-1.

Ebert, Cornelia & Christian Ebert. 2014. Gestures, demonstratives, and the attributive/referential distinction. Talk given at *Semantics and Philosophy in Europe 7*.

Ebert, Cornelia, Stefan Evert & Katharina Wilmes. 2011. Focus marking via gestures. In Ingo Reich, Eva Horch & Dennis Pauly (eds.), *Proceedings of Sinn und Bedeutung 15*, 193–208. Saarbrücken, Germany: University of Saarland.

Ebert, Cornelia, Giovanna Pirillo & Sebastian Walter. 2022. The role of gesture-speech alignment for gesture interpretation. In Sam Featherston, Robin Hörnig, Andreas Konietzko & Sophie von Wietersheim (eds.), *Proceedings of Linguistic Evidence 2020: Linguistic theory enriched by experimental data*, 65–77. Tübingen, Germany: University of Tübingen.

Engberg-Pedersen, Elisabeth. 1993. *Space in Danish Sign Language: The semantics and morphosyntax of the of space in a visual language*. Hamburg, Germany: Signum.

Goldin-Meadow, Susan. 1999. The role of gesture in communication and thinking. *Trends in Cognitive Sciences* 3(11). 419–429. 10.1016/s1364-6613(99)01397-2.

Harris, Jesse A. 2012. *Processing perspectives*. Amherst, MA: University of Massachusetts Amherst dissertation.

Harris, Jesse A. & Christopher Potts. 2009. Perspective-shifting with appositives and expressives. *Linguistics and Philosophy* 32(6). 523–552. 10.1007/s10988-010-9070-5.

Hinterwimmer, Stefan. 2017. Two kinds of perspective taking in narrative texts. In Dan Burgdorf, Jacob Collard, Sireemas Maspong & Brynhildur Stefánsdóttir (eds.), *Proccedings of Semantics and Linguistic Theory (SALT) 27*, 282–301. University Park, MD: University of Maryland. 10.3765/salt.v27i0.4153.

Hinterwimmer, Stefan. 2024. Accounts of perspective taking in narrative. *Language and Linguistics Compass* 18(3). e12517. 10.1111/lnc3.12517.

Hinterwimmer, Stefan, Umesh Patil & Cornelia Ebert. 2021. On the interaction of

gestural and linguistic perspective taking. *Frontiers in Communication* 6. 1–15. 10.3389/fcomm.2021.625757.

Kita, Sotaro & Asli Özyürek. 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 & Language* 48(1). 16–32. 10.1016/S0749-596X(02)00505-3.

Leiner, Daniel J. 2022. SoSci Survey (Version 3.3.14a). Available at https://www.soscisurvey.de.

Loehr, Daniel P. 2004. *Gesture and intonation*. Washington, DC: Georgetown University dissertation.

Maier, Emar & Markus Steinbach. 2022. Perspective shift across modalities. *Annual Review of Linguistics* 8(1). 59–76. 10.1146/annurev-linguistics-031120-021042.

McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.

Meuser, Sara. 2022. *How free is free indirect discourse? Empirical approaches to the anchoring mechanisms of perspective-taking*. Cologne, Germany: University of Cologne dissertation.

Meuser, Sara, Maximilian Hörl & Stefan Hinterwimmer. to appear. Perspective-taking and protagonist prominence: An empirical approach to the role of local and global prominence. To appear in *Discourse and Dialogue*.

Özyürek, Asli & Pamela M. Perniss. 2011. Event representations in signed languages. In *Event representations in language and cognition*, 84–107. Cambridge University Press.

Parrill, Fey. 2009. Dual viewpoint gestures. *Gesture* 9(3). 271–289. 10.1075/gest.9.3.01par.

Parrill, Fey. 2010. Viewpoint in speech-gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes* 25(5). 650–668. 10.1080/01690960903424248.

Plank, Frans. 1986. Über den Personenwechsel und den anderer deiktischer Kategorien in indirekter Rede. *Zeitschrift für germanistische Linguistik* 14(3). 284–308. 10.1515/zfgl.1986.14.3.284.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

de Ruiter, Jan P. 1998. *Gesture and speech production*. Nijmegen, The Netherlands: University of Nijmegen dissertation.

Saure, Christopher, Stefan Hinterwimmer & Anna Pia Jordan-Bertinelli. 2023. An experimental investigation of the interaction of narrators' and protagonists' perspectival prominence in narrative texts. *Zeitschrift für Sprachwissenschaft* 42(2). 341–372.

Schlenker, Philippe. 2004. Context of thought and context of utterance: A note on free indirect discourse and the historical present. *Mind & Language* 19(3). 279–304. 10.1111/j.1468-0017.2004.00259.x.

Stec, Kashmiri. 2012. Meaningful shifts: A review of viewpoint markers in co-speech gesture and sign language. *Gesture* 12(3). 327–360. 10.1075/gest.12.3.03ste.

Walter, Sebastian. 2024. The at-issue status of viewpoint gestures: Evidence for gradient at-issueness. In Geraldine Baumann, Daniel Gutzmann, Jonas Koopman, Kristina Liefke, Agata Renans & Tatjana Scheffler (eds.), *Proceedings of Sinn und Bedeutung 28*, 943–960. Bochum, Germany: Ruhr University Bochum. 10.18148/sub/2024.v28.1171.