

Multimodal Investment in Epistemic Stance Expression: Gestural and Prosodic Indicators of Speaker Commitment

Ana E. Celmare¹, Merel C.J. Scholman², Vera Demberg¹

¹Saarland University ²Utrecht University

Background. A growing body of research suggests that discourse comprehension is predictive with comprehenders continuously generating expectations about upcoming discourse using lexical, prosodic, and multi-modal information. One source of such expectations is stance-taking, which allows speakers to communicate attitudes and degrees of commitment beyond propositional content. Epistemic stance markers are particularly relevant in this regard because they signal how strongly speakers are committed to the truth of a proposition, thereby guiding expectations about the epistemic status of the discourse. Compare Examples (1) and (2): they convey the same propositional content (it is raining), but the underlined stance markers modulate how the utterances are perceived.

- (1) I'm pretty sure it's raining.
- (2) I'm certain that it's raining.

At the linguistic level, epistemic stance can be expressed through lexical means (e.g. modals, adverbs, adjectives) or discourse means (e.g. complement-taking predicates, discourse management markers) [1]. Beyond the lexico-grammatical level, gestural and prosodic resources may provide anticipatory cues that allow comprehenders to infer or predict upcoming epistemic positioning and speaker commitment, possibly even before these are fully encoded lexically [2]. For example, Figure 1 illustrates a double-handed Open Hand Supine gesture, in which the speaker raises both his arms and flips the palms to yield an upwards-oriented gestural form. This gesture has previously been identified to encode obviousness [3].

In addition to tracking epistemic certainty, comprehenders may also generate expectations about the extent to which speakers are invested in having a stance recognized or acted upon. This investment dimension shapes expectations not just about what a speaker believes, but also about the illocutionary weight they assign to their own epistemic commitment [4]. Yet, it remains unclear how multimodal cues interact with this dimension of stance. If comprehenders continuously generate predictions about speaker investment from multiple sources, then the systematic co-occurrence of gestural and prosodic cues with investment level would constitute a set of potential multimodal signals available to drive such predictions. Despite its theoretical significance, the investment dimension of stance has received limited empirical attention, and its multimodal realization remains particularly underexplored. The present study

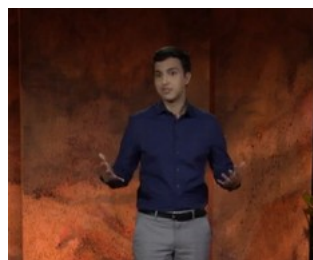


Figure 1: Illustration of gesture-stance marker co-occurrence.
S: So I'm sure he thought he was doing me a favour.
G: bimanual Open Hand Supine gesture

therefore investigates how gestural and prosodic cues contribute to comprehenders' expectations about speaker investment in expressions of high epistemic certainty.

Method & analysis. We compiled a multimodal corpus consisting of 160 video fragments from TED talk monologues [5] and spontaneous dyadic interactions [6]. This design allows for comparison of commitment marking strategies across contexts that differ in terms of planning, audience design, and interactional pressure. The fragments were selected based on the presence of one of four stance markers (*sure*, *must (have)*, *certain*, *certainly*) in the transcript. Each instance was annotated for hand gesture form, orientation, position, beat-likeness, kinesthetic features and head movement, as well as prosodic prominence operationalized as an aggregate measure of F0, energy and duration, complemented by manual ToBI annotations. Additionally, each stance expression was annotated for the speaker's degree of commitment as estimated by crowdsourced human annotators.

Preliminary annotation of the gestural dimension reveals an emerging catalogue of forms that pattern systematically with commitment level. Open Hand Supine gestures, pointing toward interlocutors, and head nods tend to co-occur with higher degrees of speaker commitment, potentially functioning as cues that raise comprehenders' expectations of strong epistemic investment. Conversely, flicks, hand cycling gestures and head tilts tend to co-occur with lower commitment, suggesting they may attenuate or revise investment expectations generated by co-occurring lexical stance markers. These patterns suggest that specific gestural form-meaning pairings function as systematic nonverbal resources for scaling epistemic investment. Prosodic and kinesthetic annotation, as well as full co-occurrence analysis across all annotation layers, are ongoing.

Implications. By decomposing the multimodal realization of epistemic stance in terms of scalar commitment, this study contributes to an understanding of how comprehenders integrate lexical, prosodic, and gestural information to generate expectations about speaker commitment and interactional intent. The findings provide insight into how local linguistic cues interact with broader multimodal signals during discourse processing, and how expectations about speaker investment contribute to coherence building and pragmatic inference.

This study is part of a larger project that will form the basis for future psycholinguistic experiments on whether specific gestures modulate perceived speaker commitment, and will provide input for computational models aimed at capturing stance and perspective-taking in multimodal communication.

References

- [1] Traugott, E. C. (2020). Expressions of stance-to-text: discourse management markers as stance markers. *Language Sciences*, 82, Article 101329. Retrived from: <https://doi.org/10.1016/j.langsci.2020.101329>
- [2] Andries, F., Meissl, K., de Vries C., Feytaerts, K., Oben, B., Sambre, P., Vermeerbergen, M., & Brône, G. (2023). Multimodal stance-taking in interaction - A systematic literature review. *Frontiers in Communication*, 8 - 2023. Article 1187977.
- [3] Jehoul, A., Brône, G. & Feytaerts, K. (2017). The shrug as marker of obviousness: Corpus evidence from Dutch face-to-face conversations. *Linguistics Vanguard*, 3(s1), 20160082. <https://doi.org/10.1515/lingvan-2016-0082>
- [4] Kiesling, S. F. (2022). Stance and stancetaking. *Annual Review of Linguistics*, 8, 409-426. <https://doi.org/10.1146/annurev-linguistics-031120121256>
- [5] Hasebe, Y. (2015). Design and Implementation of an Online Corpus of Presentation Transcripts of TED Talks. *Procedia: Social and Behavioral Sciences* 198(24), 174-182. Retrieved from: <https://yohasebe.com/tcse>
- [6] Agrawal, V., Akinyemi, A., Alvero, K., Behrooz, M., Buffalini, J., Carlucci, F. M., ... & Zollhoefer, M. (2025). Seamless interaction: Dyadic audiovisual motion modeling and large-scale dataset. *arXiv preprint arXiv:2506.22554*.