
Pathways in the signal jungle: integrating rich multimodal signals via multimodal ensembles

Petra Wagner¹, Olcay Türk¹, Hendrik Buschmeier¹
& Angela Grimmering²

¹Bielefeld University, ²Paderborn University

petra.wagner@uni-bielefeld.de, olcay.tuerk@uni-bielefeld.de,

hbuschme@uni-bielefeld.de, angela.grimmering@uni-paderborn.de

In everyday communication, speakers and listeners employ rich, multimodal communication signals which are also subject to noise and variation. This multidimensional space comprises speech (pitch movements, rhythm, tempo, voice quality, resonance characteristics) and visual signals (head movements, facial expressions, gaze, hand movements, body movements, and more). Interlocutors skillfully navigate this “signal jungle”, making informed decisions whether a signal dimension is of communicative relevance in a given moment, or not, and estimate how the relevant signal dimensions are bound into meaningful parts of a complex signal, and segregated in time. Despite this, most multimodal analyses typically limit themselves to the cross-modal coordination of a few modalities which are assumed to be crucial to constitute messages.

Building on ideas by Holler & Levinson (2019), we suggest identifying more dimensions of the rich multimodal message that are of communicative relevance in a given moment of an ongoing interaction. We argue that these constitute *multimodal ensembles* which are bound across varying signal dimensions by a set of gestalt principles: they are produced in temporal proximity, and share similar signal properties. We furthermore work on the hypothesis that prosody provides a scaffold for cross-modal integration (Wagner et al. 2014).

We exemplify our idea by a set of examples taken from dialogic interaction, and show how multimodal ensembles may signal meaning in communication. In particular, we illustrate how listeners may make use of multimodal ensembles when assessing an interlocutor’s (non-)understanding (Türk et al. 2024) or their level of cognitive load (Betz et al. 2023), and how prosodic structure may serve as a flexible cross-modal anchor. Lastly, we discuss how our ideas can be integrated in a model of multimodal interaction.

References: • Betz, S., Bryhadyr, N., Türk, O., & P. Wagner (2023). Cognitive load increases spoken and gestural hesitation frequency. *Languages*, 8(1), 71. • Holler, J., & S. C. Levinson (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, 23(8), 639–652. • Wagner, P., Malisz, Z., & S. Kopp (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209–232. • Türk, O., Lazarov, S., Wang, Y., Buschmeier, H., Grimmering, A., & P. Wagner (2024). Predictability of understanding in explanatory interactions based on multimodal cues. In *Proceedings of the 26th International Conference on Multimodal Interaction*, 449–458, Costa Rica.