MODELLING THE INTERPLAY OF SPEECH, GESTURES AND GAZE – HOW EMPIRICAL GESTURE STUDIES, EYE-TRACKING, AND INTENSIONAL LOGIC WORK TOGETHER IN RECONSTRUCTING JOINT ATTENTION AND INTENTION

We present a formal approach to support analyses and design processes of complex structures of intending in hybrid interaction scenarios between humans and machines by combining empirical gesture studies from a linguistic point of view (Bressem et al., 2013; Fricke, 2012) with experimental studies using eye-tracking data, and the formalization of complex sign processes based on intensional logic from a semiotic point of view (Posner, 1993; Siefkes, Fricke, Bressem, & Charoensit, 2023),

AUTHORS

Ellen Fricke, TU Chemnitz
Jana Bressem, TU Chemnitz
Martin Siefkes, TU Chemnitz

FUNDING

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 416228727 – SFB 1410 *"Hybrid Societies - Humans interacting with Embodied Digital Technologies"*
https://hybrid-societies.org

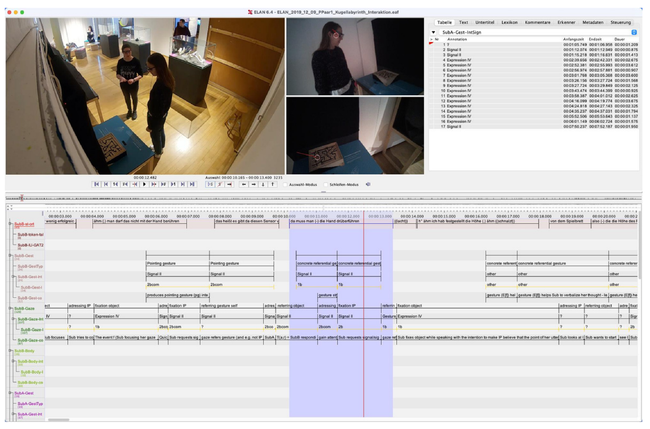
INTRODUCTION

A range of studies highlights that **intentionality and joint attention in inter-human communication are largely shaped and brought about by an interplay of verbal and bodily signs**, particularly by verbal deixis, gestural pointing, and gaze (e.g., Fricke, 2007, Tomasello, 2008, Stukenbrock, 2020). The interplay of verbal and bodily signs is also important for establishing and directing attention in human-machine encounters (e.g., Kopp, 2014; Rennert, Pfeiffer, & Wachsmuth, 2014, Staudte et al., 2011).

However, **existing studies on human-machine interaction** usually base their implementation on experimental setups **focusing on tasks with a clear goal** (e.g., sandwich making, driving a car) **with mostly one participant**. Yet, interactions in the real world involve more than one person, take place in different environments, have varying degrees of complexities, and speakers make use of more than one modality.

AIM AND METHOD

The present contribution aims to fill this gap by using **natural interaction between humans as a possible model for human-machine interaction**. The study uses **video and eye-tracking data** gathered in an experimental setting in which **15 dyads of participants** interacted with different digital exhibits in a museum. The data are annotated and analyzed in **ELAN**, combining a **cognitive-linguistic analysis** of speech-gesture relations, qualitative analyses of **interactional sequences** (Bressem, et al. 2013, Fricke, 2012, 2014, 2021), analysis of eye-tracking data for **basic eye-movement parameters** and the **formalization of complex sign processes** based on intensional logic from a semiotic point of view (Posner, 1993; Siefkes, Fricke, Bressem & Charoensit, in press).



Example of ELAN annotation

$T(b,f) \wedge I(b, E(f) \rightarrow G(a, T(b,f) \wedge I(b, E(f) \rightarrow T(a,r)))) \wedge G(b, (E(f) \rightarrow G(a, T(b,f) \wedge I(b, E(f) \rightarrow T(a, r)))) \rightarrow (E(f) \rightarrow T(a,r)))$

A person b produces the gesture f, intending the occurrence of f to make person a believe that b produces the gesture f with the intention for the occurrence of f to make person a do the reaction r of stopping, and b believes that if the occurrence of f will lead to a´s belief that b produced f with the intention for the occurrence of f to make a do r, then the occurrence of f really will make a do r.

Formalization of the occurrence of a gestures based on Posner (1993)

RESULTS

Speech, gestures, gaze indicate joint attention and intentionality along with different levels of abstraction and complexity

Distribution of **markers vary** (gaze important, bodily orientation less relevant)

Communicative **situation and task influences joint attention and gaze patterns**: even the use of verbal deictics along with a pointing gesture may be disregarded because of the more pressing task of object manipulation

Incongruence of gaze and head direction indicator for multiple attentional targets that result in particular patterns.

Formalization of sign processes from interactional data based on intensional logic possible

CONCLUSION

Communication of intentionality and establishing joint attention in (natural) interpersonal interaction is **complex** and achieved through a modality specific **interplay** of markers. A **formal description** based on Posner (1993) provides a **means to link patterns** of multimodal markers with different levels of intentionality. As such, the formalization allows for an **implementation** of these markers in human-machine communication. The results also indicate that **more research is needed that acknowledges the situational context** of establishing joint attention and intentionality in multimodal interaction, while at the same time underlining the need for **studies in more natural encounters.**

REFERENCES

Bressem, J., Ladewig, S. H., & Müller, C. (2013). Linguistic Annotation System for Gestures (LASG). In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. (Handbooks of Linguistics and Communication Science 38.1.) (pp. 1098-1125). De Gruyter Mouton. | Brône, G., & Oben, B. (Eds.). (2018). Eye-tracking in Interaction: Studies on the role of eye gaze in dialogue (Vol. 10). John Benjamins Publishing Company. | Fricke, E. (2007). Origo, Geste und Raum: Lokaldeixis im Deutschen. | Fricke, E. (2012). Grammatik multimodal: wie Wörter und Gesten zusammenwirken (Vol. 40). Walter de Gruyter. | Posner, R. (1993). Believing, causing, intending: The basis for a hierarchy of sign concepts in the reconstruction of communication. In Signs, Search and Communication (pp. 215-270). Walter de Gruyter. | Rennert, P., Pfeiffer, T., & Wachsmuth, I. (2014). Spatial references with gaze and pointing in shared space of humans and robots. In Spatial Cognition IX: International Conference, Spatial Cognition 2014, Bremen, Germany, September 15-19, 2014. Proceedings 9 (pp. 121-136). Springer International Publishing. | Siefkes, M., Fricke, E., Bressem, J., & Charoensit, A. (in press). Modelling intentional complexity in hybrid interaction scenarios beyond explicit and implicit communication. In B. Meyer, U. Thomas, & O. Kanoun (eds.), Hybrid Societies – Humans Interacting with Embodied Technologies, Vol. 1. Springer. | Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. Cognition, 120(2), 268-291. | Stukenbrock, A. (2020). Deixis, meta-perceptive gaze practices, and the interactional achievement of joint attention. Frontiers in Psychology, 11, 1779. | Tomasello, M. (2008). Origins of human communication. MIT press.