TOWARDS INTEGRATING POSE ESTIMATION WITH NEURO-IMAGING FOR THE ANALYSIS OF SIGNED LANGUAGE VIDEO STIMULI

Sébastien Vandenitte[1]
Doris Hernández[2]
Jarkko Keränen[2]
Tommi Jantunen[2]
Anna Puupponen[2]
[1]University of Namur
2University of Jyväskylä

1. BACKGROUND

- Recent technological innovation enables combining methodologies like neuroimaging and pose estimation (Trettenbrein & Zaccarella 2021, Rivolta et al. 2025).
- Ongoing study: We set out to use pose estimation in order to revisit the videos of FinSL (Finnish Sign Language) expressions used as stimuli in a neurocognitive study of enactment (Hernandez et al. 2025).
- Motivation: Hernández et al. (2025) suggest that specific attentional involvement in detection of enactment is triggered by increasing elements and saliency in signs with strong enactment vs. signs without enactment → impact of stimulus-specific kinematics
- Aim: To conduct a kinematic analysis on the video stimuli to investigate potential interactions between the EEG measurements of FinSL signers' processing of enactment and the kinematic features involved in that enactment.
- Materials: Video stimuli of isolated FinSL signs with and without enactment (Fig. 1), EEG measurements of a detection task investigating signers' recognition of enactment in signs (Fig. 2) + response-time indices of their behavioural responses to these stimuli.

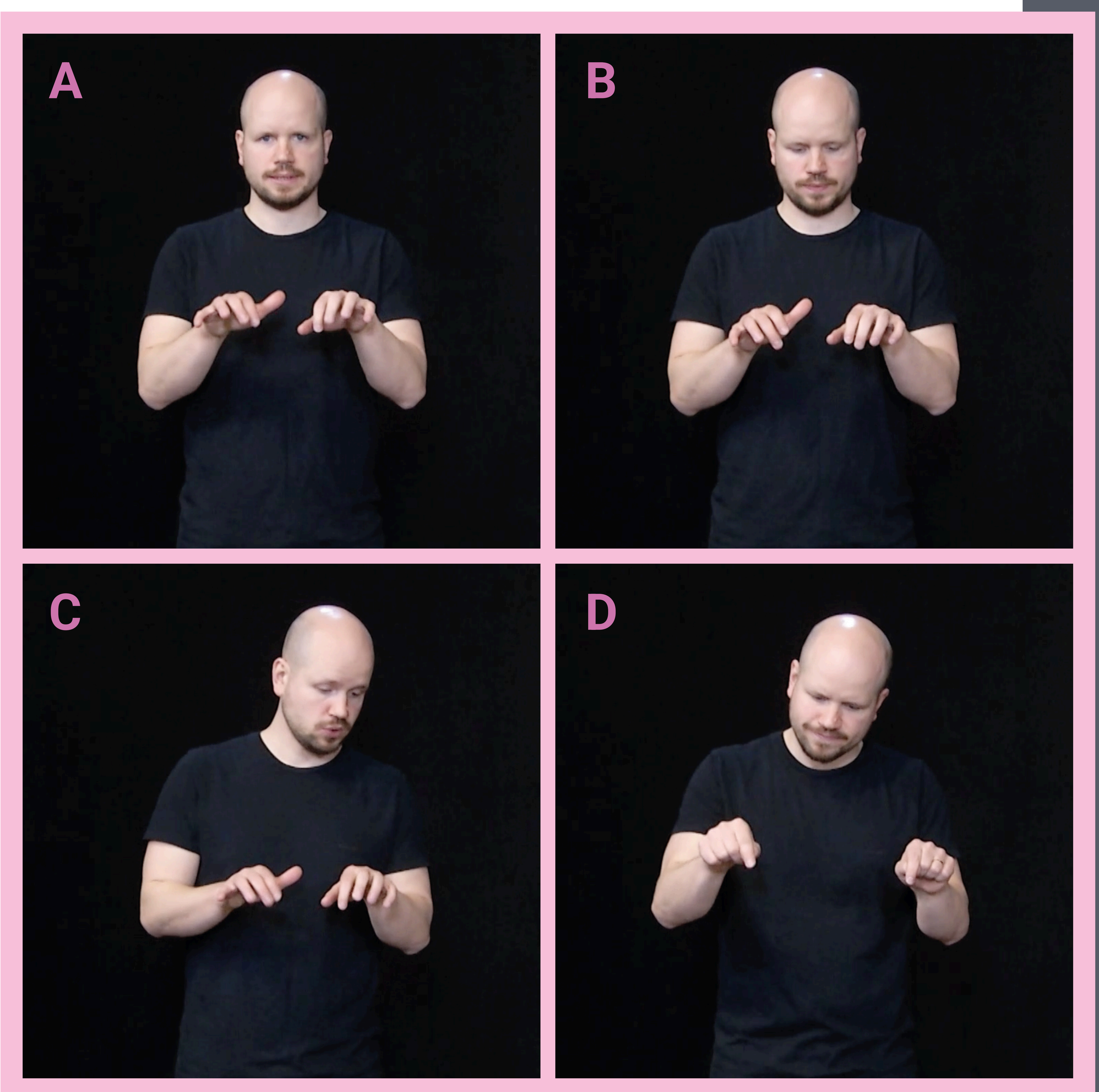

Fig.1 Stimulus set of 4 video stimuli with an increasing degree of enactment: no enactment (A), some enactment (B), substantial enactment (C), and full enactment (D).

2. NEUROIMAGING

EEG measurements:

- Continuous signal divided into epochs time-locked to the onset of each signed video (200 ms pre-stimulus baseline period prior to onset + 800 ms post-stimulus time window)
- Epoch represents a segment of the EEG signal that reflects the brain's response to a specific sign
- Analyzed in the time-domain by using average event-related potentials (ERPs) (N2, P3) for each stimulus epoch across subjects

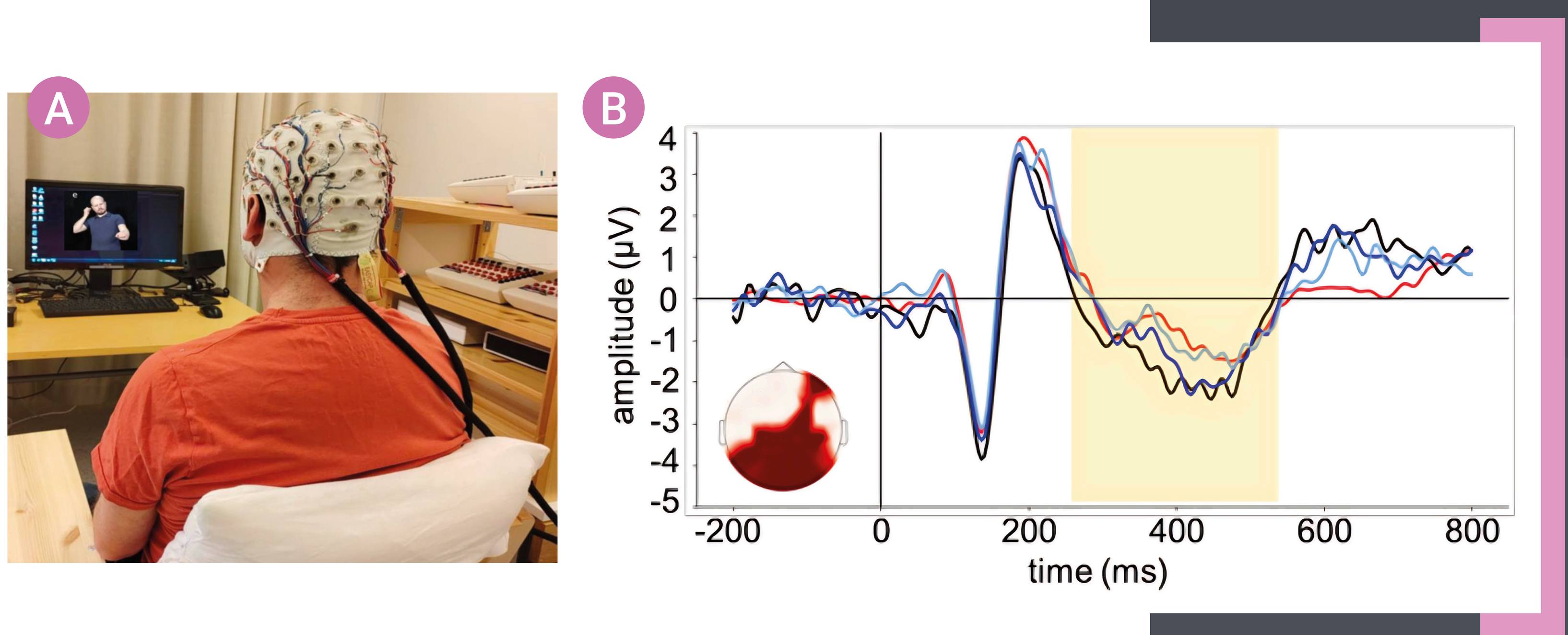

Fig.2 EEG recording setting (A). An example from Hernández et al. (2025, 145) demonstrating averaged ERPs for the amplitude contrast between signs with enactment vs. signs with no enactment (B). Yellow area indicates the time window (269-735 ms) for results associated with increased visual attention (N2pb).

3. POSE ESTIMATION

The video stimuli were processed with the MediaPipe Pose Landmarker (Lugaresi et al. 2019; Fig. 3). The x and y coordinates were used to get average kinematic values for each stimulus:

- Velocity (nose, upper torso, left wrist, right wrist)
- Acceleration (nose, upper torso, left wrist, right wrist)
- Distance between the left and right wrists
- Surface of the area delimited by the nose, the left wrist, and the right wrist

The coordinates and time derivatives were smoothed using Savitzky-Golay filters (Pouw et al. 2024)

Fig.3 Video stimulus processed for pose estimation.

4. LIMITATIONS

- Accuracy of kinematic data obtained
- Average measurements capture gross patterns
- Likely need to explore beyond the body parts and kinematics surveyed
- Other potential factors varying across videos, e.g., meaning or degree of iconicity

Finnish Cultural Foundation
Central Finland

Research Council of Finland

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

fnrs FREEDOM OF RESEARCH

UNIVERSITÉ DE NAMUR

REFERENCES

Hernández, D., Puupponen, A., Keränen, J., Vandenitte, S., Anible, B., Ortega, G. and Jantunen, T. (2025). Neuroelectrical and behavioral correlates of constructed action recognition in Finnish sign language. Neuroscience, 575, 140-149. https://doi.org/10.1016/j.neuroscience.2025.03.046.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C., Yong, M.G., Lee, J., Chang, W., Hua, W., Georg, M. and Grundmann, M. (2019). MediaPipe: A Framework for Building Perception Pipelines. ArXiv, abs/1906.08172.

Pouw, W. (2024). Wim Pouw's EnvisionBOX modules for social signal processing (Version 1.0.0) [Computer software]. https://github.com/WimPouw/envisionBOX_modulesWP.

Rivolta, C. L., Lizarazu, M., Costello, B., and Carreiras, M. (2025). Cortical tracking of sign language: The role of language knowledge in tracking of different articulators. Proceedings of the National Academy of Sciences, 122(51), e2512665122.

Trettenbrein, P. C., and Zaccarella, E. (2021). Controlling video stimuli in sign language and gesture research: The OpenPoseR package for analyzing OpenPose motion-tracking data in R. Frontiers in Psychology, 12, 628728.