To understand is to predict: machine learning identifies low-frequency entrainment to visual stimuli as the basis of sign language comprehension via predictive processing

Evie A. Malaia (University of Alabama, AL, USA), Julia Krebs (University of Salzburg, Austria), Sean Borneman (BHSS, IN, USA), Ronnie B. Wilbur (Purdue University, IN, USA)

The field of spoken language processing has established that spoken language comprehension relies on neural activity tracking entropy fluctuation in the acoustic envelope of the signal (Stilp & Kluender, 2010). Envelope tracking at a range of frequencies, termed signal-based entrainment, forms the basis of speech comprehension (Riecke et al., 2018). As compared to speech, sign language processing is not well understood. While it has been established that the visual signal for sign languages contains higher entropy than non-communicative human biological motion (Borneman et al., 2018), the question of whether human sensitivity to entropy of the signal might support sign language processing in the same manner it supports speech comprehension has not been evaluated on neurobehavioral data to date.

26-channel EEG data was recorded from 24 proficient Deaf signers, as they viewed 40 videos in Austrian Sign Language sentences (Figure 1A), pseudo-randomly mixed with time-reversed (i.e. linguistically non-interpretable) versions of the same videos (as well as fillers). Behavioral data confirmed that the participants did not consider time-reversed videos of sign language sentences linguistically acceptable (M=1.72, SD=.76, where response 1 denoted stimuli that were not acceptable as sign language, and 7 denoted understandable sign language on a 7-point Likert scale; we take the low rating as evidence that the videos were generally not understandable). Sign language videos were rated as linguistically acceptable (M=5.8, SD=1.05).

Optical flow, characterizing distribution of velocities in an image, was calculated for each pixel of the stimulus videos using MATLAB Vision toolbox (Figure 1B-C). Coherence between optical flow in the stimulus and EEG neural response (per video, per participant) was then computed using canonical component analysis with MATLAB NoiseTools toolbox. Peak correlations were extracted for each frequency for each electrode, participant, and video, and used as input into machine learning classification pipeline (data set consisted of 24 participants x 2 states as instances, and 62 coherence peak metrics binned by .2 Hz ranges as parameters).

During classification, two ensembles (AdaBoost and Random Forest) achieved 100 percent out-of-sample prediction accuracy on hold-out dataset for the whole brain. To quantify the contribution of frequency-specific coherence measures to successful classification, we repeated the analysis, while reducing the number of input parameters used as input to 5 parameter vectors at a time: e.g. vectors for frequencies 0.2 to 1 Hz (0.2, 0.4, 0.6, 0.8, and 1.0). The results are summarized in Table 1. Notice that lower frequency ranges, .2-1 Hz and 3.2-4 Hz, yield higher out-of-sample prediction accuracy, with classification accuracy of 100 percent attained for Extra Trees classifier. This is notable because of how this reflects on the data used for training the classifier. Decision tree algorithms are sensitive to the specific data on which they are trained; tree-based boosting aggregation ensemble algorithms work better if the predictions from the submodels (based on a limited number of features) are uncorrelated or weakly correlated. Additionally, using the best features of the data for training can significantly boost performance of tree-based algorithms.

The findings demonstrate that cortical tracking of entropy in the visual signal of sign language relies on lower (under 4 Hz) frequencies, despite high spatiotemporal frequencies inherent in sign language signal (Borneman et al., 2018). This indicates that online comprehension is likely supported by predictive processing mechanisms grounded in existing sign language knowledge, and driven by low-frequency (temporal and spatial) visual signal.

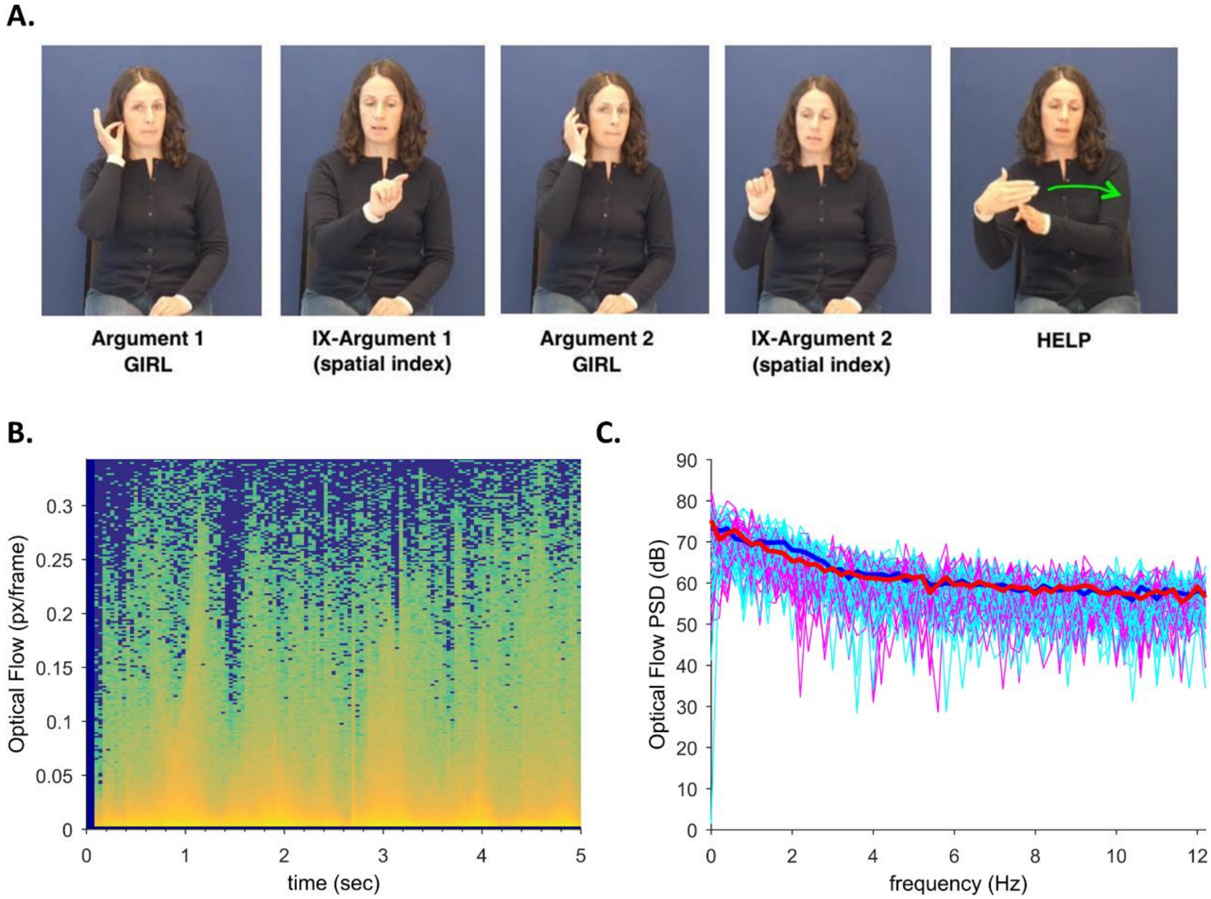

Figure 1. A. Dynamic signed sentence in Austrian Sign Language as a sequence of still frames. B. Optical flow data in time domain for one sign language video. C. Comparison of PSD of optical flow in frequency domain for sign language (magenta) and time-reversed videos (blue).

Table 1: Accuracy (*M, SD*) of classification for comprehension from EEG frequency subsets

EEG frequency subset	AdaBoost	Gradient Boosting Machine	Random Forest	ExtraTrees Classifier
0.2 – 12.4 Hz	97 (.06)	95 (.04)	99 (.02)	97 (.04)
0.2 – 1.0 Hz	99 (.03)	97 (.03)	99 (.03)	100 (.03)
3.2 – 4.0 Hz	78 (.15)	84 (.10)	81 (.13)	70 (.07)
9.2 – 10.0 Hz	63 (.07)	68 (.11)	61 (.08)	53 (.11)

References:

Borneman, J. D., Malaia, E. A., & Wilbur, R. B. (2018). Motion characterization using optical flow and fractal complexity. *Journal of Electronic Imaging, 27*(05), 051229.

Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2018). Neural entrainment to speech modulates speech intelligibility. *Current Biology, 28*(2), 161–169.

Stilp, C. E., & Kluender, K. R. (2010). Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. *Proceedings of the National Academy of Sciences, 107*(27), 12387–12392.