Low-Frequency Entrainment to Visual Motion Underlies Sign Language Comprehension

E. A. Malaia, S. C. Borneman, J. Krebs, and R. B. Wilbur

***Abstract*— When people listen to speech, neural activity tracks the entropy fluctuation in the acoustic envelope of the signal. This signal-based entrainment has been shown to be the basis of speech parsing and comprehension. In this electroencephalography (EEG) study, we compute sign language users' cortical tracking of changes in visual dynamics of the communicative signal in the time-direct videos of sign language, and their time-reversed counterparts, and assess the relative contribution of response frequencies between.2 and 12.4 Hz to comprehension using a machine learning approach to brain state classification. Lower frequencies of EEG response (.2-4 Hz) yield 100% classification accuracy, while information about cortical tracking of the visual envelope in higher frequencies is less informative. This suggests that signers rely on lower visual frequency data, such as envelope of visual signal, for sign language comprehension. In the context of real-time language processing, given the speed of comprehension responses, this suggests that fluent signers employ a predictive processing heuristic based on sign language knowledge.**

***Index Terms*— EEG, sign language, perceptual sampling, vision, language comprehension.**

I. Introduction

THE field of spoken language processing has accumulated substantial correlational evidence that spoken language comprehension relies on neural activity tracking entropy fluctuation in the acoustic envelope of the signal [1]–[3]. This envelope tracking at a range of frequencies between 100 Hz and 1 kHz (i.e. matching the human vocal range, cf. [4]), also sometimes termed signal-based entrainment, or frequency-following response (FFR), forms the basis of speech parsing and comprehension [5], [6]. As compared to speech, sign language processing and comprehension is not well understood, and lacks neurocomputational processing models. While it has been established that the visual signal for sign languages contains higher entropy (i.e. are less predictable across multiple time-scales) than non-communicative human biological motion [7]–[9], the question of whether human sensitivity to entropy of the visual signal might support sign language processing in the same manner it supports speech comprehension has not been posited to date. An EEG study of signers' and non-signers entrainment to the amplitude of visual frequencies in sign language (quantified by Instantaneous Visual Change (IVC) metric, equivalent to loudness for speech) indicated that in frontal regions, fluent signers showed stronger coherence to IVC than non-signers [10]. Low-frequency entrainment to sign language video signal 'loudness' was found in both signers and non-signers between 0.4 and 5 Hz, peaking at 1 Hz. However, lack of a baseline condition - a non-linguistic visual stimulus – prevents a conclusive interpretation of the results, which are also counter-intuitive in light of current understanding that visual cortex responds in the alpha band in response to aperiodic stimulation [11], [12]. Thus, if modality-driven preferences determine the spectrum of entrainment for the stimuli, then peak coherence to sign language visual stimuli in both signers and non-signers should be observed around alpha frequency (8-12 Hz). It has thus remained unclear whether similarity between signers' and non-signers' neural responses performance reflected sign language processing per se (which non-signers did not know), or was part of the response to the lower-level visual features associated with the IVC metric. The work to understand neural bases for scene categorization, on the other hand, has identified links between neural activity and visual stimuli, separating the timecourses for visual feature encoding (i.e. bottom-up processing, occurring at 90 ms post-stimulus onset, or 10 Hz), as well as higher-level cognitive processing, such as categorization (or top-down processing, peaking between 150 and 200 ms after image onset and persisting across the trial epoch, below 5 Hz [13]). This suggests that an entrainment to a signal above 10 Hz (or in the range of alpha frequency in EEG response) is likely to be elicited by low-level (higher-frequency) processing of rapidly changing visual features. On the other hand, top-down processing based on global scene categorization (or lexical retrieval, in case of sign language) would be expected to yield a lower-frequency response (under 5 Hz). To investigate signers' response to multiple visual frequencies in the visual signal, we designed an experiment to assess the relative contribution

Manuscript received June 3, 2021; revised October 10, 2021 and November 4, 2021; accepted November 7, 2021. Date of publication November 11, 2021; date of current version December 2, 2021. This work was supported in part by the National Science Foundation under Grant 1932547 and Grant 1734938 and in part by the National Institute of Health (NIH) (R01) under Grant 108306. *(Corresponding author: E. A. Malaia.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by The University of Salzburg Institutional Review Board under Protocol No. EK-GZ: 07/2018.

E. A. Malaia and S. C. Borneman are with the Department of Communication Disorders, The University of Alabama, Tuscaloosa, AL 35487 USA (e-mail: eamalaia@ua.edu).

J. Krebs is with the Center for Cognitive Neuroscience, Department of Linguistics, University of Salzburg, 5020 Salzburg, Austria (e-mail: julia.krebs@sbg.ac.at).

R. B. Wilbur is with the Department of Speech, Language, and Hearing Sciences and the Department of Linguistics, Purdue University, West Lafayette, IN 47907 USA (e-mail: wilbur@purdue.edu).

Digital Object Identifier 10.1109/TNSRE.2021.3127724

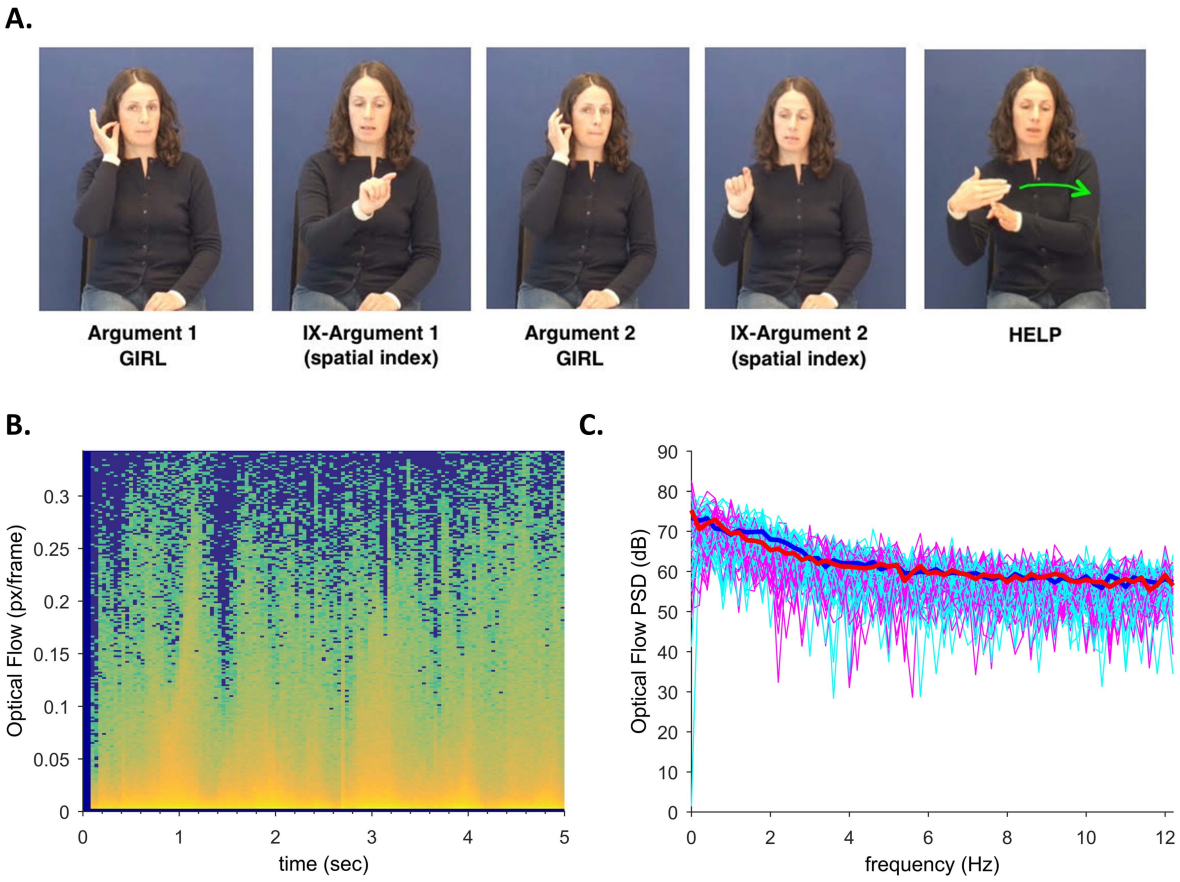

Fig. 1. a) a dynamic signed sentence as a sequence of still frames; b) optical flow data in time domain; c) comparison of PSD of optical flow in frequency domain for sign language (magenta) and reversed videos (blue).

of frequencies from.2 to 12.5 Hz to the measures of cortical coherence to changes in the signal. We hypothesized a range of possible outcomes for the investigation. Based on prior research, we hypothesized a range of possible outcomes for the investigation. Prior observations of entrainment to visual stimuli in the alpha (8-12 Hz) range [14] might suggest that sign language, as aperiodic stimulus, is processed in the bottom-up manner, based on rapid dynamic changes in the visual signal. Alternatively, if signers' knowledge of the sign language allow for lower-frequency sampling of the visual input, and reliance on predictive processing during language comprehension [13], then observations of lower (under 5 Hz) frequencies of EEG exhibiting coherence with sign language and reversed videos stimuli would be expected.

II. Materials and Methods

A. Participants

Proficient users of Austrian Sign Language (ÖGS) were recruited, who reported normal or corrected-to-normal vision, and no history of neurological disorders. All of the participants used ÖGS as their primary language in daily life, and were members of the Deaf community in Austria. Their sign language proficiency was tested by a certified sign language interpreter. 24 participants (13 male) aged between 28 and 68 years (M = 42, SD = 12.27) took part in the study. All procedures in the study were undertaken with the understanding and written consent of each subject. The study conforms to the Declaration of Helsinki (World Medical Association, 2013). The Institutional Review Board of the University of Salzburg approved the design of the study and engagement of human participants.

B. Stimuli and Procedures

Each participant was shown a mixed set of videos, which contained 40 videos that were sentences in Austrian Sign Language (ÖGS, signed by a fluent signer), and 40 videos which were time-reversed versions of these sentences (i.e. not linguistically acceptable), as well as filler videos. Filler videos consisted of videos of sign language sentences with classifier constructions and topicalized sentences using SOV and OSV word order, as well as simple sentences with SOV word order (200 total filler sentences). Neural and behavioral responses to filler videos used to prevent habituation are reported in detail in [15], [16]. Overall, behavioral data in response to filler stimuli were similar to that for time-direct videos under analysis. However, as spectro-temporal parameters of these videos differed from the ones considered here, they were not amenable to the same type of analysis. Sign language stimuli consisted of dynamic videos of

signed sentences, easily understood by proficient sign language users (see Figure 1). The list of glossed sentence translations is provided in the appendix. To produce linguistically non-acceptable stimuli, we time-reversed sign language videos, to hold constant the spatiotemporal frequencies of the visual stimuli in sign-language and non-sign language categories. Time-reversed videos thus contained no comprehensible sign language; we also obtained behavioral responses for each stimulus, evaluating proficient signers' assessment of how linguistically acceptable each stimulus was (a single stimulus contained a video of a signed sentence, or a time-reversed representation of it). The conditions were pseudo-randomized (such that no condition repeated more than twice in a row). Two different pseudo-random orders of stimuli were used, balanced among participants. Each participant was presented with a training block of videos prior to the experiment, to become familiar with task requirements, and to ask any questions they had. The videos were presented on the screen 35.3 x 20 cm in size. The size of the videos was 1280 x 720 pixels. Participants were asked to avoid excessive motion during the presentation of the video material. Every trial began with the presentation of a fixation cross (2000 ms) to allow the participant to prepare; this was followed by a 200 ms presentation of an empty black screen, and then the stimulus video, which appeared in the middle of the screen. At the end of each trial, a question mark appeared in the center of the screen for 3000 ms, during which the participants were instructed to perform the rating task by pressing a key on the keyboard. In the rating task, participants had to rate the videos on a scale from 1 to 7 (1 for 'that is not ÖGS'; 7 for 'that is good ÖGS', and 4: not ÖGS, but understandable).

C. EEG Data Acquisition and Processing

Data collection was carried out on a 26-channel EEG system at a rate of 500 Hz using active electrodes. The electrodes were placed on the participant's scalp according to the standards of the 10/20 system (Fz, Cz, Pz, Oz, F3/4, F7/8, FC1/2, FC5/6, T7/8, C3/4, CP1/2, CP5/6, P3/7, P4/8, O1/2), and secured with an elastic cap (Easy Cap, Herrsching-Breitbrunn, Germany). The impedances of all electrodes were kept below 5 kΩ. The eye movements and blinks were monitored and recorded using electrodes placed over the right and left outer canthi (horizontal eye movement, HEOG), and left inferior and superior orbital ridge (vertical eye movement, VEOG).The AFz electrode functioned as the ground electrode during the recording. All electrodes were referenced to the electrode on the left mastoid bone. At the start of each trial, numerical trigger codes were sent by the stimulus presentation computer to the EEG recording computer, and time-stamped on the EEG recordings for synchronization. Offline, following the recording, electrodes were re-referenced to the averaged data from the electrodes at the left and right mastoids. The signal was filtered with a bandpass filter (Butterworth Zero Phase Filters; high pass: 0.1 Hz, 48 dB/Oct; low pass: 30 Hz, 48 dB/Oct) in Brain Analyzer. As we planned to analyze coherence of video and EEG data, the higher frequency data (EEG) needed to be downsampled. As video data was recorded at 25 fps, the highest computable frequency for it – Nyquist frequency – is 12.5 Hz. The signal was then corrected for ocular artifacts using the Gratton and Coles method [17], and segmented from recorded triggers – the onsets of video stimuli to 5 seconds following the onset. The full duration of video stimuli was between 5 to 7 seconds; the 5 second cutoff ensured that only neural responses to ongoing video stimuli were analyzed (see Figure 2).

D. Optical Flow Extraction From Video Stimuli and Coherence Calculation

Optical flow is a technique frequently utilized in computer vision to quantify the motion of image content between two adjacent frames of a video recording. Optical flow is a metric that tracks signal variability across time by quantifying the velocity magnitude of each object (based primarily on edge contrast values) in pixels per frame. Although optical flow analysis converts each frame to a velocity profile, it does not filter the spatial content of dimensions, as the resulting signal contains velocity per pixel versus time, preserving both the spatial and temporal information available in the video. Based on optical flow, the velocity signal is analyzed according to fractal complexity using the formula $M(f) = \frac{\alpha}{f^{\beta}}$, where M is the power spectral density profile of the signal (PSD), f is the frequency, α is the PSD magnitude, and β is the parameter for fractal complexity of the signal. By computing the optical flow measure in video, we quantify the distance traveled by each individual pixel as it moves from frame to frame, such that intensity of optical flow is proportional to the area of the moving part in the video. Optical flow was computed for each stimulus video using the vision toolbox optical flow function from MathWorks' MATLAB. This function produces an output matrix of size equal to the input video frame, such that each element of the matrix identifies the magnitude of optical flow velocity (pixels per frame) between the two frames for each corresponding pixel in the video. An optical flow histogram (which can be thought of as a velocity spectrum) is thus created for each frame of the stimulus video. Then, for each frame (25 frames for each second of the video), the amplitudes across all velocity bins were added to calculate the total magnitude of optical flow for each frame, which was used an instantaneous measure of motion in the stimuli. Across multiple frames this produced an optical flow timeseries.

Coherence between the optical flow timeseries of each stimulus video and the neural response timeseries in each electrode for each participant was then calculated. To compute coherence at a given frequency, both timeseries were first filtered at that frequency (from 0.02 Hz to 12.5 Hz, as limited by the 25 fps video frequency) using a second-order IIR bandpass filter. The filtered timeseries correlation was then calculated using canonical correlation analysis with MATLAB NoiseTools toolbox [18]. Both the peak correlation and the timeshift of that correlation were extracted for each frequency for each participant, stimulus video, and electrode location.

E. Data Setup and Pipelines for Machine Learning

Our intent for machine learning analysis was, first, to assess predictability of the two conditions, time-direct sign language

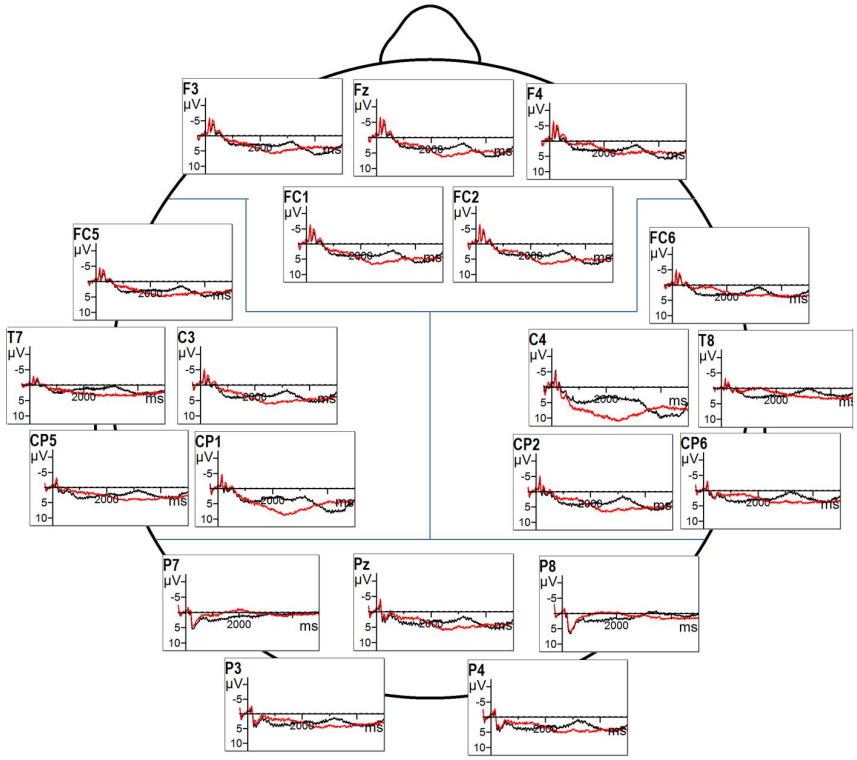

Fig. 2. Comparison of EEG responses to sign language (black) and time-reversed videos **(red)**. For the purposes of presenting the data, ERPs are baseline-corrected using 300 ms epoch prior to each trigger; negative is plotted upward. Blue lines indicate electrode clusters used for analysis (anterior (FC1, FC2, F3, F4, Fz); posterior (P3, P4, P7, P8, Pz); left (FC5, C3, CP1, CP5, T7); right (FC6, C4, CP2, CP6, T8).

and time-reversed sign language stimuli, from the neural data on frequency coherence with the stimuli videos' optical flow. The secondary goal was to evaluate the predictive value of input parameters – in this case, frequency bins of coherence data – for such classification. To construct the data matrix, we used the peak cross-correlation values from 62 frequency ranges (from 0.2 Hz to 12.4 Hz in 0.2 Hz increments) over each of the four brain regions (anterior, comprising data from electrodes in positions F3, F4, Fz, FC1, and FC2; posterior comprising data from electrodes in positions P7, P8, P3, P4, Pz; left, including data from electrodes C3, FC5, T7, CP1, CP5; and right, with the data from C4, FC6, T8, CP2, CP6) for each of the 0.2 Hz-wide frequency bins of optical flow PSD, and each participant. As differing data distribution can negatively impact performance of machine learning algorithms by over-weighting less clustered input parameters, we performed scaling data transform such that each of the parameters would have a mean value of zero and a standard deviation of one. Six classifier algorithms were used to evaluate the performance: two linear algorithms (Linear Regression (LR) and Linear Discriminant Analysis (LDA)), and four nonlinear algorithms (k-nearest neighbors (kNN), classification and regression trees (CART), Naïve Bayes (NB), and support vector machines (SVM)), with default tuning parameters of Python *sklearn* library. Machine learning algorithms, in general, are data-greedy methods that create complex representation models based on raw data; however, the algorithms vary in terms of weighting of different parameters of the raw data; thus, it is rarely possible to determine in advance, which types of algorithms will perform well on the data. The six classifier algorithms chosen included a variety of algorithms differing in assumptions about the data. For example, linear algorithms (Linear Regression (LR) and Linear Discriminant Analysis (LDA) assume Gaussian distribution of the data, but differ in terms of performance on well-separated classes (i.e. LR can be unstable, while LDA is more appropriate). Among the four nonlinear algorithms used, classification and regression trees (CART) are simple self-correcting (pruning) algorithms that perform well in the presence of outliers. Naïve Bayes (NB) algorithm, on the other hand, assumes conditionally independent parameters (i.e. non-interacting ones) – a very strong assumption, which rarely holds on real data; the algorithm, nevertheless, can perform well on data sets where parameter dependence is noisy. K-nearest neighbors (kNN) algorithm makes no assumptions about the functional form of the classification problem, but, on the other hand, is highly reliant on training data, such that it performs well in situations where training and testing data sets

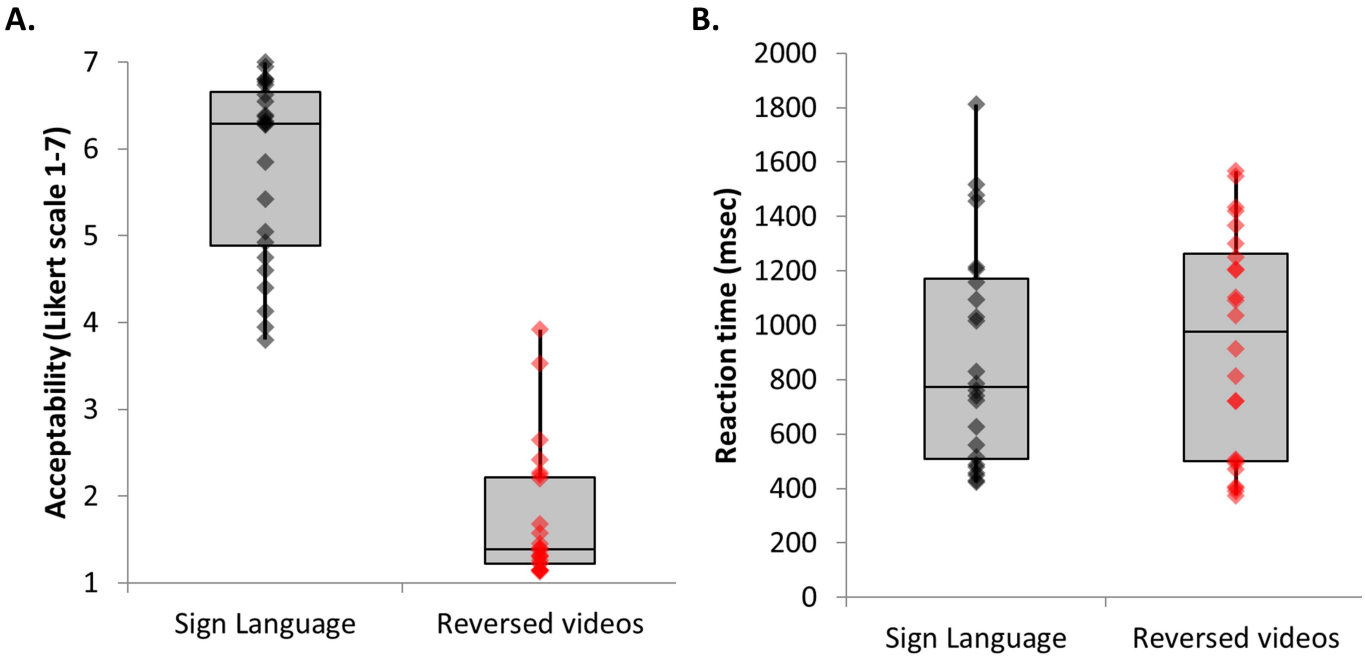

Fig. 3. Behavioral response distribution in Sign Language and Reversed Videos conditions. A. Acceptability ratings on Likert scale (7 – good Austrian Sign Language; 4 - not Austrian Sign Language, but understandable; 1 - not Austrian Sign Language), with significantly lower ratings for time-reversed videos; B. Reaction times (in ms) to the two conditions did not differ significantly.

are very similar (i.e. individual participants' parameters are alike across the population). Support vector machines (SVM) are flexible in terms of analysis, and can learn problem representation from the data itself, but are, as a result, the most data-greedy among the options. Application of multiple algorithms to the data set in its entirety, as well as sub-sets, provides the most thorough understanding of potential models that can best describe the data. During classification, 20% of the data was retained for a validation hold-out set (sample of data held back from the rest of the analysis and modeling). We used a 10-fold cross-validation approach with the test harness pipeline configuration to prevent data leakage between training and testing data in each cross-validation harness.

III. Results

A. Behavioral Results

Data from the participants' behavioral responses on the Likert scale from 7 ('that is good Austrian Sign Language') through 4 ('not Austrian Sign Language, but understandable'), to 1 ('that is not Austrian Sign Language') indicated that only sentences in the sign language condition were rated as linguistically acceptable, while reversed videos of sign language (i.e. not linguistically acceptable videos) were not considered meaningful communication (sign language $M = 5.80$; $SD = 1.48$; reversed videos $M = 1.72$, $SD = 1.35$) (see Figure 3). Paired t-test between individual ratings of time-direct and time-reversed videos indicate significantly higher ratings ($t(23) = 14.01$; $p < .001$) for time-direct videos. Response times did not differ significantly between conditions ($t(23) = -1.3$; $p > .2$; sign language $M = 883$ ms; $SD = 535$ ms; reversed videos $M = 925$ ms, $SD = 541$ ms).

B. Machine Learning Results

Peak coherence between the stimuli and neural activity occurred between 100 ms and 250 ms post-stimulus onset in response to both time-direct and time-reversed (not linguistically acceptable) video stimuli conditions, as expected for visual dynamic stimuli (cf. [13]). The cross-correlation matrix of the input vectors (frequency coherence bins between EEG and optical flow in the visual stimuli) to machine learning pipeline is presented in Figure 4. The red line along the diagonal represents self-correlation of individual input parameters (coherence frequency bins). Notice the structure in the matrix around the diagonal in the quadrant encompassing.4 to 4 Hz bins: dark blue suggests high values of negative correlation of the parameters and red indicates high positive correlation values, both of which are likely to weigh strongly in classification. We used the bagged decision tree classifier (*ExtraTreesClassifier* from Python *sklearn* library) to estimate the importance of input parameters (i.e. coherence frequencies) in the data set. The importance scores for all input parameters was <0.01, with the exception of frequency bins 0.8 Hz (importance score 0.16), and 1Hz (importance score 0.16), highlighting relevance of these input parameters for classification accuracy. The accuracy metrics for the algorithms are summarized in Table I. Five algorithms (LR, LDA, kNN, Naïve Bayes, and SVM) achieved 100% out-of-sample prediction accuracy on hold-out dataset for the whole brain data set. For region-specific analysis classification accuracy remained above 80% accuracy, with most values above 90% accuracy. Nonlinear Naïve Bayes performed at 100% accuracy for each brain region separately as well as for the whole-brain dataset.

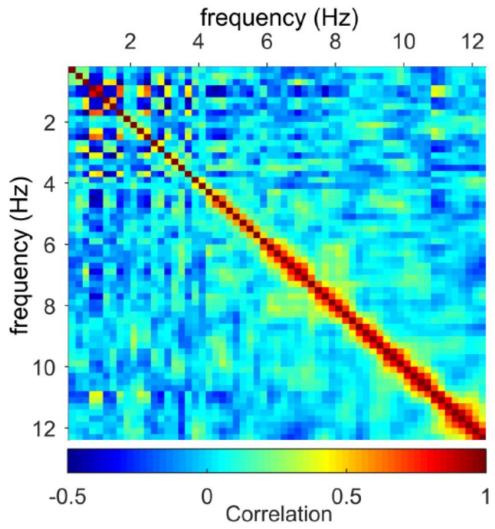

Fig. 4. Cross-correlation matrix of the input vectors (coherence between EEG and optical flow in the visual stimuli, binned in.2 Hz frequency increments). Both horizontal and vertical axes represent the same bins (top to bottom and left to right). The red line along the diagonal represents self-correlation of individual input parameters (value of 1, or perfect self-correlation).

TABLE I
ACCURACY (IN PERCENT) ON HOLD-OUT SET FOR WHOLE-BRAIN, AND SPATIALLY LOCALIZED INPUT PARAMETERS ACROSS CLASSIFICATION ALGORITHMS. NOTICE HIGH PERFORMANCE OF LINEAR AND NON-LINEAR ALGORITHMS ACROSS BRAIN REGIONS

Accuracy	LR	LDA	kNN	CART	NB	SVM
Whole brain	100	100	100	96	100	100
Anterior	98	100	95	90	100	98
Posterior	100	89	98	100	100	100
Left	100	93	100	80	100	100
Right	100	97	97	90	100	100

The high prediction accuracy in classifying the time-direct vs. time-reversed conditions suggests that machine learning algorithms successfully identify common neural responses to visual linguistic stimuli based on stimulus-EEG coherence data across frequency bins. To identify the independent contribution of each frequency to successful classification, we repeated the analysis, this time reducing the number of input parameters used 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 II. Notice that lower frequency ranges (up to 4 Hz) yield highest out-of-sample prediction accuracy, with classification accuracy of 100% attained for a set of frequency bins between.2 and 1 Hz using kNN, NB, and SVM algorithm, while higher frequencies appear to contribute less to recognition of neural state for language comprehension.

IV. DISCUSSION

To assess the relationship between neural data response to visual entropy of the signal in linguistic (sign language) vs. not linguistically acceptable (time-reversed sign language) conditions, electroencephalography signal (EEG) was recorded from participants who were fluent signers, while they were viewing

TABLE II
ACCURACY (IN PERCENT) ON HOLD-OUT SET FOR SPECIFIC FREQUENCY BINS OF COHERENCE PARAMETERS. NOTICE >80% ACCURACY IN IDENTIFICATION OF STIMULI TYPE AND COMPREHENSIBILITY ACROSS ALGORITHM TYPES FOR LOW-FREQUENCY (UP TO 4 Hz) DATA

Feature bins	LR	LDA	kNN	CART	NB	SVM
0.2 - 1.0 Hz	99	99	100	94	100	100
1.2 - 2.0 Hz	92	93	89	86	90	93
2.2 - 3.0 Hz	93	93	90	83	91	93
3.2 - 4.0 Hz	82	83	71	75	81	88
4.2 - 5.0 Hz	67	67	62	61	69	73
5.2 - 6.0 Hz	61	61	55	57	59	61
6.2 - 7.0 Hz	79	78	79	70	67	79
7.2 - 8.0 Hz	66	66	52	58	56	60
8.2 - 9.0 Hz	51	51	49	58	43	50
9.2 - 10.0 Hz	57	60	52	64	56	52
10.2 - 11.0 Hz	78	86	54	71	67	67
11.2 - 12.0 Hz	63	64	50	52	54	58

sign language sentence videos, and the same videos that were time-reversed. The participants rated the sentences on a Likert scale from 1 ("that is not Austrian Sign Language") to 7 ("that is good Austrian Sign Language"). The sign language videos and time-reversed videos differed only in the time direction of the signal; all other spectro-temporal parameters of the videos were the same. To relate the neural data to the video data (sign language signal and time-reversed sign language stimuli), we first quantified the video signal using changes of optical flow across multiple visual frequencies. This measure was linearly regressed against individual EEG signals of each participant, such that peak cross-correlation frequency was defined for each channel in the EEG data. We then employed a variety of machine learning pipelines to evaluate whether brain state of processing sign language was classifiable from the state of watching time-reversed videos, equivalent in low-level features; we also assessed relative contribution of brain regions and specific frequencies to classification accuracy of six machine learning algorithms. What we probed in our study is whether neural response to the motion frequencies of the signal is based on low-level feature assessment (in this case, low-level, or sensory features describe high frequency motion at the onset and offset of the signs), or on assessment of whether the visual data contains any vocabulary items of the sign language known. If high-frequency data were predictive of comprehension, it would indicate that low-level (motion-based) features are indeed critical for sign recognition. However, as we identified low-frequency data as predictive of comprehension, this suggests that participants rely not on local (higher-frequency) motion features, but rather on slower (lower-frequency) global visual features. The only possibility to do real-time processing and respond rapidly to comprehension questions would be to employ predictive processing: i.e. using low-frequency sampling of the input signal, to retrieve a number of potentially appropriate lexical/syntactic items from sign language vocabulary, and rapidly reject those that do not fit the signal at the next data point. The results indicated that electrophysiological responses to visual language yield enough information for successful classification. Cross-correlation analysis indicated that frequencies under 4 Hz tended to contribute most weight

to classification accuracy. Feature evaluation using the bagged decision tree classifier also highlighted importance of 0.8 Hz and 1Hz input parameters to classification accuracy. In general, frequency-based entrainment to stimuli is an overarching cortical mechanism of sensory processing evidenced across modalities. For spoken language (with high temporal variability of the signal) it is the envelope features, which describe entropy fluctuations of the changing signal, that predict comprehensibility of the signal [1], [2], [19]. For sign language, circumstantial evidence from multiple behavioral studies has pointed to the likelihood of a similar mechanism. For example, [20] investigated the ability of signers and non-signers (native users of Spoken English/American Sign Language and Spoken Chinese/Chinese Sign language, respectively) to parse dynamic point-light presentations of 'pseudo-hieroglyphic writing' (with novel stimuli created for the experiment to mitigate for the Chinese Sign Language users' familiarity with hieroglyphic Chinese) [20]. Native users of either sign language were able to perceptually separate discrete segments, such as 'strokes' and 'transitions', in the signal, while non-signers perceived point-light motion as continuous. Since no linguistic cues were available to signers in these studies, motion entropy envelope tracking might be one perceptual adaptation enhanced in sign language that might allow signers to parse dynamic visual stimuli relying on language-driven skills.

Infant studies provide further evidence for attentional relevance of entropy-rich portions of visual signal during sensitive period for language acquisition. [21] investigated infants' attentiveness to fingerspelled stimuli in sign language. The operational definition of visual sonority of the stimuli used in the study is qualitatively based on Brentari's [22] model of sign language phonology, and is functionally equivalent to an entropy measure in visual modality. In a preferential looking paradigm, hearing 6-month-olds looked significantly longer at high-entropy stimuli than low-entropy stimuli. This preference disappeared in older infants (around 12 months of age) who had not received any signed language experience. Perceptual sensitivity to entropy (syllabicity) to auditory stimuli is known to peak around six months and specialize to environmental input by approximately 12 months of age [23], [24].

The present study links the research on neurobiologically-motivated approaches to language comprehension [1] and action processing [25] to computational modeling of information transfer in communication [9], [26]. Both behavioral (acceptability) and neural measures of comprehension of sign language sentences appear rooted in entrainment of neural activity to the dynamic variations in the entropy of the visual signal, as measured by optical flow. The findings demonstrate that cortical tracking of spectro-temporal dynamic entropy in the visual signal of sign language relies on lower (under 4 Hz) frequencies, and is likely mediated by predictive processing mechanisms based on language knowledge. Identification of common mechanisms underlying comprehension in speech and sign, based on neural response to linguistic stimuli that tracks the low-frequency envelope of signal entropy, could help develop brain-based diagnostics for language processing disorders for users of sign languages.

There were several limitations to the present analysis which related primarily to the scope of the chosen question. We did not specifically address the relationship between behavior (and between-participant behavioral variability) and the neural signal. In sign language populations, the participant pool is often heterogenous, as strict inclusion criteria (participant age, etc.) would severely limit the pool of participants. Language proficiency assessment tools, or detailed description of the grammatical structure, are yet unavailable for ÖGS. Additionally, analysis focused on frequency-domain correlation between the signal and the neural response, rather than the relative location or timing of entrainment, or possible variations among participants –questions which certainly deserve further scrutiny. It is possible that individual signs, particularly those symmetric in the time domain, could have been understood by signers. However, most signs (especially verb signs [27]) follow a non-symmetric dynamic trajectory, with higher acceleration at the end of the sign. Time-reversed sentence-level stimuli, thus, violated both syntactic (word order) and phonological (motion profile) rules of sign language, and, as such, were rated below understandable threshold. The stimuli in the present study were controlled for syntactic structure (simple sentences) to avoid possible variability in processing strategies that are often seen in the processing of more complex sentences [28]. The question of possible variability in processing of more complex sentences is very interesting, and should be subject to further research.

Acknowledgment

The authors would like to thank all Deaf informants taking part in the present study. They would also like to thank Waltraud Unterasinger for signing the stimulus material.

References

[1] C. E. Stilp and K. R. Kluender, "Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 27, pp. 12387–12392, Jul. 2010.

[2] C. E. Stilp and K. R. Kluender, "Stimulus statistics change sounds from near-indiscriminable to hyperdiscriminable," *PLoS ONE*, vol. 11, no. 8, Aug. 2016, Art. no. e0161001.

[3] J. E. Peelle, J. Gross, and M. H. Davis, "Phase-locked responses to speech in human auditory cortex are enhanced during comprehension," *Cerebral Cortex*, vol. 23, no. 6, pp. 1378–1387, Jun. 2013.

[4] N. Guo *et al.*, "Speech frequency-following response in human auditory cortex is more than a simple tracking," *NeuroImage*, vol. 226, Feb. 2021, Art. no. 117545.

[5] M. P. Broderick, A. J. Anderson, G. M. di Liberto, M. J. Crosse, and E. C. Lalor, "Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech," *Current Biol.*, vol. 28, no. 5, pp. 803–809, 2018.

[6] N. Ding, L. Melloni, H. Zhang, X. Tian, and D. Poeppel, "Cortical tracking of hierarchical linguistic structures in connected speech," *Nature Neurosci.*, vol. 19, no. 1, pp. 158–164, 2016.

[7] E. Malaia, J. D. Borneman, and R. B. Wilbur, "Information transfer capacity of articulators in American sign language," *Lang. Speech*, vol. 61, no. 1, pp. 97–112, Mar. 2018.

[8] S. Z. Gurbuz *et al.*, "A linguistic perspective on radar micro-Doppler analysis of American sign language," in *Proc. IEEE Int. Radar Conf. (RADAR)*, Apr. 2020, pp. 232–237.

[9] J. D. Borneman, E. Malaia, and R. B. Wilbur, "Motion characterization using optical flow and fractal complexity," *J. Electron. Imag.*, vol. 27, no. 5, p. 1, Jul. 2018.

[10] G. Brookshire, J. Lu, H. C. Nusbaum, S. Goldin-Meadow, and D. Casasanto, "Visual cortex entrains to sign language," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 24, pp. 6352–6357, Jun. 2017.

[11] M. Senoussi, J. C. Moreland, N. A. Busch, and L. Dugué, "Attention explores space periodically at the theta frequency," *J. Vis.*, vol. 19, no. 5, p. 22, May 2019.
[12] C. S. Y. Benwell, C. Keitel, M. Harvey, J. Gross, and G. Thut, "Trial-by-trial co-variation of pre-stimulus EEG alpha power and visuospatial bias reflects a mixture of stochastic and deterministic effects," *Eur. J. Neurosci.*, vol. 48, no. 7, pp. 2566–2584, Oct. 2018.
[13] M. R. Greene and B. C. Hansen, "Disentangling the independent contributions of visual and conceptual features to the spatiotemporal dynamics of scene categorization," *J. Neurosci.*, vol. 40, no. 27, pp. 5283–5299, Jul. 2020.
[14] K. E. Mathewson, C. Prudhomme, M. Fabiani, D. M. Beck, A. Lleras, and G. Gratton, "Making waves in the stream of consciousness: Entraining oscillations in EEG alpha and fluctuations in visual awareness with rhythmic visual stimulation," *J. Cognit. Neurosci.*, vol. 24, no. 12, pp. 2321–2333, Dec. 2012.
[15] J. Krebs, E. Malaia, R. B. Wilbur, and D. Roehm, "Interaction between topic marking and subject preference strategy in sign language processing," *Lang., Cognition Neurosci.*, vol. 35, no. 4, pp. 466–484, May 2020.
[16] J. Krebs, E. Malaia, R. B. Wilbur, and D. Roehm, "Psycholinguistic mechanisms of classifier processing in sign language," *J. Experim. Psychol., Learn., Memory, Cognition*, vol. 47, no. 6, pp. 998–1011, Jun. 2021.
[17] G. Gratton, M. Coles, and E. Donchin, "A new method for off-line removal of ocular artifact," *Electroencephalogr. Clin. Neurophysiol.*, vol. 55, no. 4, pp. 468–484, 1983.
[18] A. de Cheveigné, D. D. E. Wong, G. M. Di Liberto, J. Hjortkjaer, M. Slaney, and E. Lalor, "Decoding the auditory brain with canonical component analysis," *NeuroImage*, vol. 172, pp. 206–216, May 2018.
[19] H. R. Bosker and O. Ghitza, "Entrained theta oscillations guide perception of subsequent speech: Behavioural evidence from rate normalisation," *Lang., Cognition Neurosci.*, vol. 33, no. 8, pp. 955–967, Sep. 2018.
[20] E. Klima, O. Tzeng, Y. Fok, U. Bellugi, D. Corina, and J. Bettger, "From sign to script: Effects of linguistic experience on perceptual categorization," *J. Chin. Linguistics Monograph Ser.*, pp. 96–129, Jan. 1999.
[21] A. Stone, L.-A. Petitto, and R. Bosworth, "Visual sonority modulates infants' attraction to sign language," *Lang. Learn. Develop.*, vol. 14, no. 2, pp. 130–148, Apr. 2018.
[22] D. Brentari, *A Prosodic Model of Sign Language Phonology*. Cambridge, MA, USA: MIT Press, 1998.
[23] K. Byers-Heinlein and C. T. Fennell, "Perceptual narrowing in the context of increased variation: Insights from bilingual infants," *Develop. Psychobiol.*, vol. 56, no. 2, pp. 274–291, Feb. 2014.
[24] D. Maurer and J. F. Werker, "Perceptual narrowing during infancy: A comparison of language and faces," *Develop. Psychobiol.*, vol. 56, no. 2, pp. 154–178, Feb. 2014.
[25] A. Blumenthal-Dramé and E. Malaia, "Shared neural and cognitive mechanisms in action and language: The multiscale information transfer framework," *WIREs Cognit. Sci.*, vol. 10, no. 2, Mar. 2019, Art. no. e1484.
[26] B. J. West, E. L. Geneston, and P. Grigolini, "Maximizing information exchange between complex networks," *Phys. Rep.*, vol. 468, nos. 1–3, pp. 1–99, Oct. 2008.
[27] J. Krebs *et al.*, "Event visibility in sign language motion: Evidence from Austrian sign language (ÖGS)," in *Proc. Annu. Meeting Cognit. Sci. Soc.*, vol. 43, no. 43, 2021, pp. 362–368.
[28] E. Malaia, R. B. Wilbur, and C. Weber-Fox, "ERP evidence for telicity effects on syntactic processing in garden-path sentences," *Brain Lang.*, vol. 108, no. 3, pp. 145–158, Mar. 2009.