



Predicting naming responses based on pre-articulatory electrical activity in individuals with aphasia



Janina Wilmskoetter^{a,*}, John Del Gaizo^{b,1}, Lorelei Phillip^c, Roozbeh Behroozmand^c, Ezequiel Gleichgerrcht^a, Julius Fridriksson^c, Elynn Riley^d, Leonardo Bonilha^a

^a Department of Neurology, Medical University of South Carolina, SC, USA

^b Biomedical Data Science and Informatics, Medical University of South Carolina, SC, USA

^c Department of Communication Sciences and Disorders, University of South Carolina, SC, USA

^d Department of Communication Sciences and Disorders, Syracuse University, NY, USA

ARTICLE INFO

Article history:

Accepted 14 August 2019

Available online 26 August 2019

Keywords:

Aphasia

Stroke

EEG

Naming

Machine learning

HIGHLIGHTS

- Pre-articulatory neural activity can predict correct naming responses in individuals with aphasia.
- Electrode/time-range/energy combinations with the highest accuracies varied between individuals.
- Future individualized pre-articulatory models could be used to predict and treat aphasic utterances.

ABSTRACT

Objective: To investigate whether pre-articulatory neural activity could be used to predict correct vs. incorrect naming responses in individuals with post-stroke aphasia.

Methods: We collected 64-channel high density electroencephalography (hdEEG) data from 5 individuals with chronic post-stroke aphasia (2 female/3 male, median age: 54 years) during naming of 80 concrete images. We applied machine learning on continuous wavelet transformed hdEEG data separately for alpha and beta energy bands (200 ms pre-stimulus to 1500 ms post-stimulus, but before articulation), and determined whether electrode/time-range/energy (ETE) combinations were predictive of correct vs incorrect responses for each participant.

Results: The five participants correctly named between 30% and 70% of the 80 stimuli correctly. We observed that pre-articulatory scalp EEG ETE combinations could predict correct vs incorrect responses with accuracies ranging from 63% to 80%. For all but one participant, the prediction accuracies were statistically better than chance.

Conclusions: Our findings indicate that pre-articulatory neural activity may be used to predict correct vs incorrect naming responses for some individuals with aphasia.

Significance: The individualized pre-articulatory neural pattern associated with correct naming responses could be used to both predict naming problems in aphasia and lead to the development of brain stimulation strategies for treatment.

© 2019 International Federation of Clinical Neurophysiology. Published by Elsevier B.V. All rights reserved.

1. Introduction

Virtually all individuals with stroke-related language impairment (aphasia) regardless of the aphasia type, present with word-finding deficits. As such, individuals with aphasia make

mistakes when attempting to name objects. These naming errors are typically unpredictable, not item-specific and often vary across testing sessions (Howard et al., 1984; Freed et al., 1996). The reason for these behaviors is not well understood. However, identifying the underlying neurophysiological processes may be helpful to develop and apply individualized interventions for treating anomia in patients with aphasia.

Naming is a multi-step process starting with visual recognition followed by activation of conceptual-semantic and phonological features, which are then followed by articulation of phonetic units

* Corresponding author at: Department of Neurology, Medical University of South Carolina, 96 Jonathan Lucas St., Charleston, SC 29466, USA.

E-mail address: wilmskoe@musc.edu (J. Wilmskoetter).

¹ Equal contribution.

(Humphreys et al., 1999; Nozari et al., 2010; Gleichgerrcht et al., 2015). These steps occur within a few hundred milliseconds after an individual sees an image or object and can be delayed in individuals with aphasia. Research suggests that visual processing occurs approximately within the first 0 to 200 ms, semantic processing within 100 to 300 ms, and phonological processing within 250 to 450 ms after stimulus presentation (Vihla et al., 2006; Laganaro et al., 2009; Laganaro et al., 2013; Singh et al., 2018). Successful completion of this rapid process is orchestrated by a complex and distributed neural network, which is primarily lateralized to the left hemisphere (DeLeon et al., 2007; Indefrey, 2011; Baldo et al., 2013). Incorrect naming responses can result from errors during any stage of this multi-step process. These errors likely relate to the insufficient emergence of appropriate patterns of neural activity. In individuals with aphasia, it has been shown that pre-articulatory electrical activity, as measured with electroencephalography (EEG), differs between correct and incorrect responses (Riley et al., 2017; Singh et al., 2018). Depending on the size and location of the stroke lesion, recovery of naming deficits often engages perilesional brain regions (Meinzer et al., 2004; Lidzba et al., 2012).

Our previous research in individuals with aphasia suggests that pre-articulation electrical activity differs between correct vs. incorrect responses (Riley et al., 2017; Singh et al., 2018). If it were possible to accurately predict a patient's response before the patient initiates articulation, proactive interventions could be implemented to modify the patient's naming response. For example, the possibility of improving an individual's performance between the presentation of a target stimulus and the individual's response has been successfully shown in a recent study using EEG-based closed-loop neurostimulation to improve memory (Ezzyat et al., 2018). In order to implement a similar application for naming abilities in individuals with aphasia, the underlying neurophysiological properties of correct and erroneous responses need to be established together with developing a method using these properties to accurately predict responses that require intervention. Our study aimed to build and expand on findings of previous research by using methods that 1) allow a detailed analysis of neurophysiological electrical activity during naming, and 2) test if they can be used to accurately predict erroneous naming attempts. Specifically, we applied machine learning to predict accuracy of object naming from high-density EEG (hdEEG). The model was trained and tested on different electrode/time-range/energy-band (ETE) combinations of the input EEG data. We hypothesized that left perilesional brain regions would show distinct neural activation patterns for correct compared to incorrect naming of objects. Due to

delays and variation in semantic and phonological processing of individuals with aphasia, we hypothesized that the distinct neural patterns would become apparent between 200 and 800 ms after the presentation of the object.

2. Materials and methods

2.1. Participants

We retrospectively evaluated EEG and behavioral data from the first five consecutive individuals with chronic post-stroke aphasia who were recruited by an on-going larger study conducted in the Aphasia Lab at the University of South Carolina (Table 1). As part of the larger study, all participants evaluated in this study completed EEG and behavioral testing. Participants did not receive speech and language treatment at the time of the EEG testing. Inclusion criteria of the larger study were: 1) Left hemisphere ischemic stroke confirmed by MRI or CT scan; 2) Aphasia diagnosis; 3) At least six months post-stroke; 4) Monolingual native English speaker; 5) Ages 30 to 80 years; 6) Able to provide verbal or written informed consent. Exclusion criteria were self-reported history of 1) Dementia, traumatic brain injury, or psychiatric disorder; and 2) Alcohol abuse.

Aphasia type and severity were determined with the Western Aphasia Battery (revised version) (Kertesz, 2007). Three participants were classified with Broca's, one participant with anomic, and one participant with conduction aphasia. All but one participant showed evidence for apraxia of speech, and two participants showed evidence for dysarthria as estimated by the apraxia of speech rating scale (ASRS) (Strand et al., 2014).

The stroke lesion was located in the left hemisphere in every participant (Fig. 1). All participants were right-hand dominant before their stroke and developed aphasia following acute vascular insult. Thus, we assumed that the left hemisphere was language-dominant before their stroke.

2.2. Naming task

Each participant performed a naming task while hdEEG recordings were obtained. The naming task consisted of 120 stimuli that were drawings of 80 concrete and 40 abstract images (Fig. 2). For the study presented here, we only included the concrete images into our analyses. Of note, the abstract images were used for other contrasts, which were not included in this study. Participants were instructed to name all concrete images and to remain silent for

Table 1
Participant characteristics. WAB-AQ = Aphasia Quotient of the Western Aphasia Battery (revised version).

Variables	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
Demographics					
Age (years)	54	59	42	54	54
Sex	Male	Male	Female	Female	Male
Race	White	Black	White	Black	White
Stroke					
Years since stroke	5	8	10	10	1.5
Lesion volume (ml)	240	103	59	146	55
Aphasia					
Aphasia type	Broca's	Anomic	Broca's	Broca's	Conduction
WAB-AQ (max. 100)	57.5	83.2	55.2	74.8	51.7
Co-occurring Speech Impairments*					
Apraxia of Speech (max. 4)	3	3	4	3	0
Dysarthria (max. 4)	0	3	2	0	0

* Apraxia of speech and dysarthria severity were estimated each with a 5-point Likert scale as described in the apraxia of speech rating scale (ASRS) (Strand et al., 2014) with the following score definitions: 0 = not present; 1 = detectable but infrequent; 2 = frequent but not pervasive; 3 = nearly always evident but not marked in severity; 4 = nearly always evident and marked in severity.

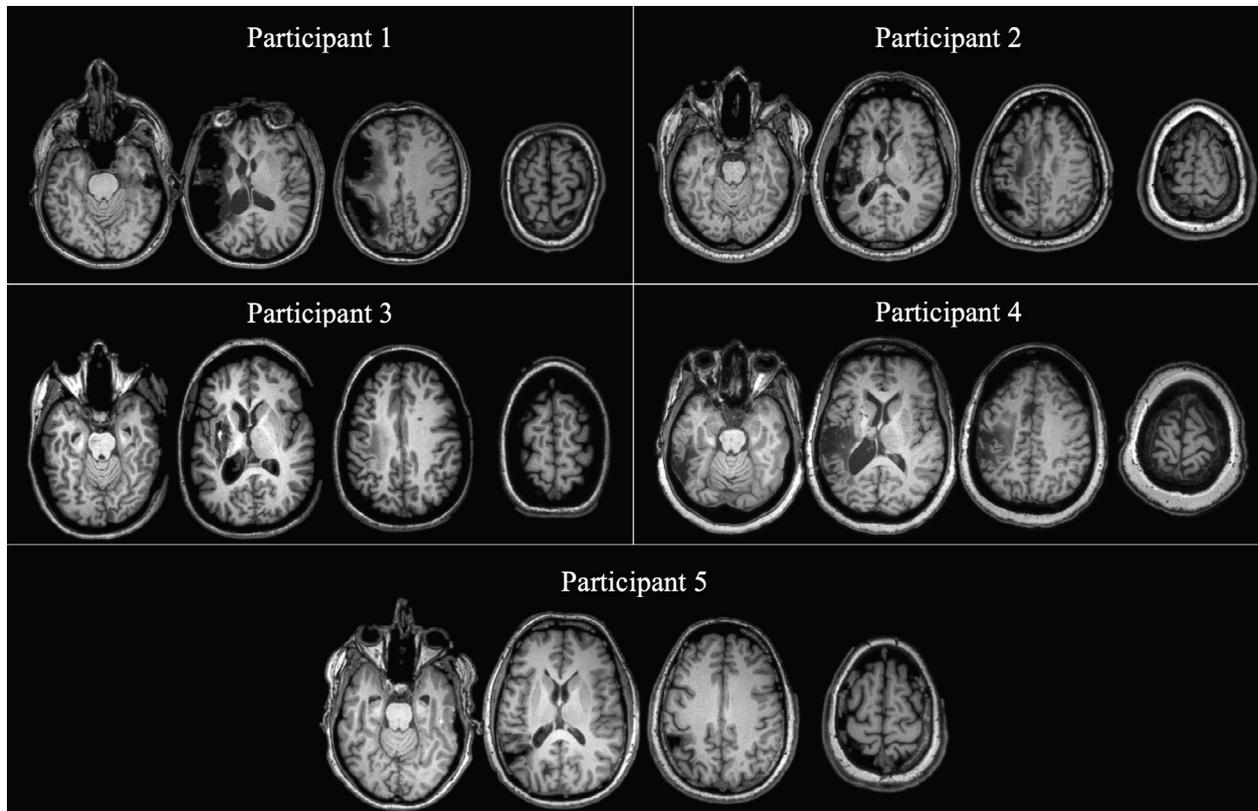


Fig. 1. T1- weighted MR Images from each participant (left side = left hemisphere, right side = right hemisphere).



Fig. 2. Naming tasks consisted of the presentation of in total 80 concrete (e.g. deer, cookies) and 40 abstract images, with 8 seconds between images.

every abstract image. Images were presented for 8 seconds each, and both image types were intermixed. Prior to beginning of the task, participants were trained on the naming task. They demonstrated understanding by responding appropriately to both image types on a total of five consecutive trials.

The entire naming task was audio-recorded and later transcribed to classify responses as correct or incorrect. The onset of the stimuli presentations was time locked with a brief audio signal, which facilitated measurement of the duration between the onset of the stimulus presentation and the onset of the participant's response. If patients did not respond to a stimulus, we excluded these instances from calculating average response latencies. We used PRAAT (version 6.0.37) (Boersma et al., 2018) for transcription and response latency measures.

For the study presented here, we only analyzed participants' responses to concrete stimuli. Responses were classified as correct if the first naming attempt was either the target word ("cookie(s)" for "cookies"; "deer" for "deer") or a consistent subordinate semantic word of the target (e.g., "chocolate cookies" for "cookies"; "stag" for "deer"). Phonetic (articulatory) distortions – as they would occur as evidence of apraxia of speech or dysarthria – counted as correct. Responses were classified as incorrect if any other utterance was produced as the first naming attempt, e.g. phonemic or semantic errors, superordinate semantic errors (e.g.,

"food" for "cookie"; "animal" for "deer"), neologisms, unrelated responses, or circumlocutions. No responses were also classified as incorrect.

2.3. EEG acquisition

Neural activity was recorded with 64-channel high density scalp EEG using the Brain Vision active electrode system (Brain Products GmbH, Germany) placed on a standard electrode cap (Easy-Cap GmbH, Germany). The electrode placement on the cap followed the 64 electrode layout that includes the electrodes from the 10–20 system (<https://www.easycap.de/wordpress/>) (Singh et al., 2018; Seo et al., 2019). The EEG signals were recorded using a common reference. A BrainVision actiCHamp amplifier (Brain Products GmbH, Germany) on a computer utilizing Pycorder software was used to record the EEG signals at 1 kHz sampling rate after applying a low-pass anti-aliasing filter with 200 Hz cut-off frequency.

2.4. EEG data processing

We processed the hEEG data for every participant using the software CURRY (version 8, Compumedics Neuroscan, Germany). First, we applied a common average reference to the EEG signals,

followed by a band-pass filter (1–70 Hz) and notch filter (60 Hz, slope 1.5 Hz). We excluded unusually noisy channels that we detected by visual inspection. Artifacts resulting from eye-blink and associated muscle activities were identified in the prefrontal channels (FP1 And FP2) as epochs exceeding 60 μV . We averaged the eye-blink artifacts and conducted a principle component analysis. Removal of the first principal component resulted in removal of the eye-blink artifacts with minimal impact on the remaining hdEEG signals. Epochs in any channel exceeding $\pm 350 \mu\text{V}$ were classified as bad blocks and excluded from further analyses.

Using the pre-processed EEG data, we calculated stimulus-locked event-related potentials (ERPs) of pre-articulatory brain activity for every stimulus. Each ERP epoch had a total duration of 1700 ms spanning from 200 ms before stimulus presentation (used as baseline) to 1500 ms after stimulus presentation. We chose this time window according to previous research associating language processing during naming with early time windows after stimulus presentation (Laganaro et al., 2009; Laganaro et al., 2011; Singh et al., 2018). Further, we sought to choose a time window before the onset of response articulation to avoid interference of muscle activities with the EEG signal. Hardly any responses across the five participants were articulated within 1000 ms post-stimulus presentation (6 out of 397 = 0.02%), thus we focused our analyses on effects occurring before 1000 ms.

2.5. Machine learning

We provide the MATLAB code used for this project together with a tutorial in the [supplementary material](#).

2.5.1. Data dimensions

We decided a priori to restrict the analysis for each participant to the same set of electrodes that measure EEG signal from frontal, temporal, and parietal electrodes. Based on previous research on the importance of left hemisphere areas for language performance and aphasia recovery (DeLeon et al., 2007; Fridriksson et al., 2010; Indefrey, 2011; Fridriksson et al., 2012; Baldo et al., 2013), we included electrodes in the left hemisphere only.

There were 79–80 concrete trials per subject. Each trial was sampled for 1700 ms, with the first 200 ms sampled before the stimuli was shown to the participant. The samples started at time 0, and at a rate of 1 kHz for a total of 1701 time samples per trial.

2.5.2. EEG wavelet transform

The data were input to a continuous wavelet transform (CWT) (MATLAB Wavelet Toolbox, function “cwt” defaults of Morse wavelet, γ equal to 3, time-bandwidth product set to 60, and 10 voices per octave) between the frequencies of 8 and 30 Hz. With these parameters, the “cwt” function outputs 32 values at each time point, one for each of the frequency values between 8 and 30 such that the values in the frequency array halve every 10th element; for example, the 1st element was 30, the 11th element was 15, the 21st element was 7.5, etc. Therefore, for each trial, the output at this step was a 3D CWT magnitude matrix of size number of electrodes by 1701 (1700 ms sampled at 1 kHz) by 32.

The next step was to separate the magnitudes of the CWT matrix into the energy bands of alpha and beta; which we defined at 8–12 Hz and 13–30 Hz, respectively. This resulted in 2 separate matrices for the 2 energy bands.

Each of the frequency-bands had defined time buffers as well, where samples were discarded that were within the buffer from either the start or the end of the signal. The buffers were 200 ms for alpha, and 100 ms for beta. These values were set based on the time resolution of the corresponding energy-band's CWT.

The time resolution of a Morse CWT is inversely proportional to the time duration of the Gaussian envelope. The envelope duration

decreases with wavelet frequency and increases with time-bandwidth product. Therefore, lower frequencies lead to longer envelope duration and lower time resolution given constant time-bandwidth product. If the sample time duration is too short, most of the envelope energy will not be within the sample. For example, the sample duration has to be at least 200 ms to capture about 50% of the energy of a Morse wavelet with a frequency of 8 Hz and the aforementioned CWT parameters. In this scenario, approximately 50% of the Gaussian window's signal energy is outside the 200 ms window.

Other windowing options have pros and cons. The Gaussian envelope time-duration is proportional to the square root of the time-bandwidth product. Changing the time-bandwidth product from 60 to 20 would narrow the envelope enough in time so that 75% of the signal energy would be calculated within the baseline 200 ms. A con is that the decreased envelope time-duration would lower the frequency resolution.

For the first 200 ms of the recording (–200 to 0), no stimuli were presented, but only a fraction of it could be used depending on the frequency band's time resolution. No baseline was used for alpha as the pre-stimuli signal does not remain after accounting for the 200 ms buffer. However, we used the 100 points before the presentation as a baseline for the beta energy. The baseline energy was calculated as the mean over the pre-stimuli time points. The correction was simply a subtraction of this mean from the rest of the signal.

Finally, each of the 2 energy matrices were averaged over the frequencies at each time point to obtain 2 energy vectors for each trial. To reduce computation, we down sampled by a factor of 5.

In summary, each alpha vector consisted of the averaged CWT energy between 8 and 12 Hz, sampled every 5 ms between 200 and 1300 ms post stimulus. Each beta vector consisted of the averaged CWT energy between 13 and 30 Hz, sampled every 5 ms between 100 and 1400 ms post stimulus. Therefore, a length 220 ($\frac{1300-200}{5}$) alpha vector and a length 260 ($\frac{1400-100}{5}$) beta vector was produced for each trial. Only the beta vector was processed with baseline noise correction.

2.5.3. Classification

The classification pipeline was used to determine which electrode/time-range/energy (ETE) combinations were most predictive of whether or not the participant correctly named the concrete object. For each subject, classification was performed on a per-trial basis (i.e., to determine the neural signature that accurately classified performance on the most trials for that individual). Therefore, the classifier used at a particular ETE for one participant could not be used for a different participant. We used leave-one-out cross validation (LOOCV) to measure the pipeline's accuracy on unseen data, while maintaining a maximum number of training points. For example, if a subject's data consisted of 80 concrete trials, 79 of these trials were used for the training set and 1 trial was used for the test set on each iteration of the cross validation.

The first step was to convert the signal power (energy) within different frequency bands (i.e. Alpha and Beta) into scores (Fig. 3). At each ETE point, the median of the correct responses and incorrect responses were calculated. From these medians, a score was calculated with the below formula:

$$\text{score}(x) = x - \text{median}_{\text{correct}} + x - \text{median}_{\text{incorrect}}$$

$$\therefore \text{score}(x) = 2x - (\text{median}_{\text{correct}} + \text{median}_{\text{incorrect}})$$

where x is the magnitude of the energy signal at the given ETE.

If $\text{median}_{\text{correct}} > \text{median}_{\text{incorrect}}$ and $x > \text{median}_{\text{correct}}$, then the score will be positive. If $\text{median}_{\text{correct}} > \text{median}_{\text{incorrect}}$ and $x < \text{median}_{\text{incorrect}}$, then the score will be negative. If

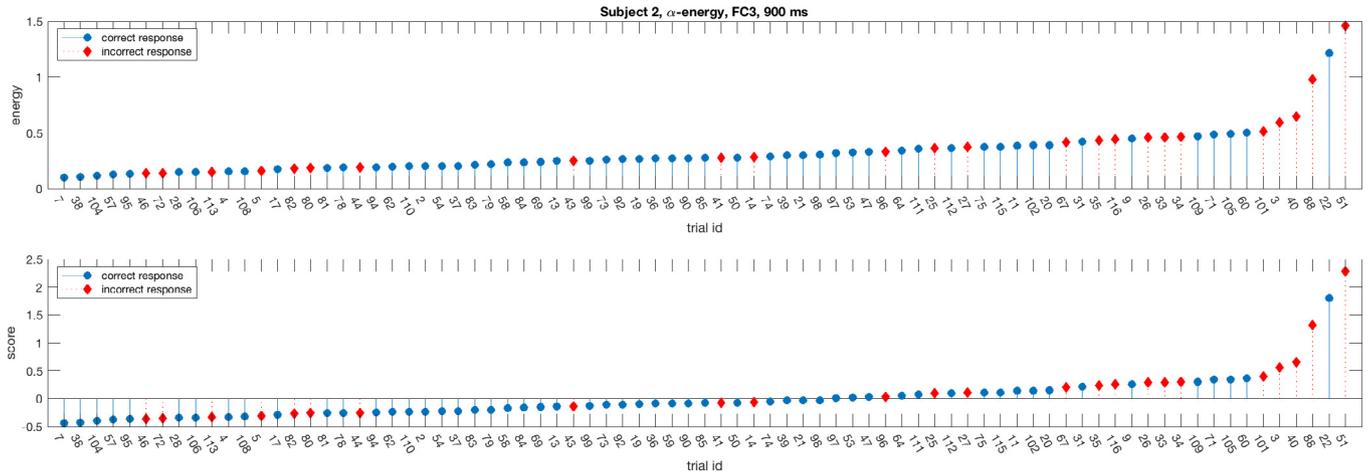


Fig. 3. A score is calculated for each trial at each time point and continuous wavelet transform (CWT) magnitude for a given subject: $score(x) = 2x - (median_{correct} + median_{incorrect})$. In the pictured example (participant 2, electrode FC3, 900 ms, alpha-energy), correct (incorrect) responses mostly correspond to negative (positive) scores. For most of the machine learning analyses, only the training data is used to calculate $median_{correct}$ and $median_{incorrect}$ to prevent overfitting. For permutation analysis, the entire data set is used to calculate the medians. However, the entire dataset is also used to calculate the medians for each of the permutations. Therefore, the accuracy metric calculated on the real data does not have an artificial boost when compared to the null distribution of accuracies. Note: only trials of concrete images are shown, however the IDs extend to 120 due to the summative number of concrete and abstract images.

$median_{correct} > x > median_{incorrect}$, then the score will either be positive if x is closer to $median_{correct}$; or negative if x is closer to $median_{incorrect}$. If $median_{correct} < median_{incorrect}$, then the results will be similar except the sign of the score will be negated.

We did not use means as the median is more robust to noise. To prevent overfitting, the medians were calculated with the training data only on each run of the LOOCV.

A probit regression model was fit to the training set scores. We did not calculate a bias term in this step. Instead, the previous score calculation was meant to center the values near 0 (Fig. 3). In a probit regression model, the coefficients can be interpreted as z-scores. The coefficients were found by iteratively maximizing the below log-likelihood:

$$LL = \sum_{i=1}^{n_{trials}} y_i \ln(\Phi(score_i \beta)) + (1 - y_i) \ln(1 - \Phi(score_i \beta))$$

where $\Phi(x)$ equals the probability that $x > 0$, which is the probability that $score_i$ times the β coefficient was a correct trial; and y_i is whether trial i was correct (1), or incorrect (0). If the scores tend to be negative for correct responses, then the β coefficient will also be negative.

Finally, we fit and applied the score calculation and probit regression model n times, where n is the number of trials, for each of the n folds in the LOOCV. The accuracy was calculated between the n predictions and n true responses.

2.5.4. Permutation analysis

The classification pipeline was also used for permutation analysis. However, it was not computationally feasible to perform LOOCV for each permutation. Instead, the score and probit pipeline was trained and tested on the entire data set at each ETE point. The accuracy was calculated from the predictions of the entire set. We then compared this accuracy with a null distribution.

To generate the null distribution, a normal distribution was fit to the accuracies resulting from employing the classification pipeline to 50 permutations at each ETE point. At each permutation, the incorrect/correct trials were randomly shuffled. The p-value was calculated as the probability of obtaining the true-accuracy given this null distribution, single tailed.

We heuristically determined 50 permutations. The shape of the random distribution was readily apparent by this point. More

permutations produced little information gain for the intensive computation.

The p-values were corrected using false discovery rate (FDR) (Benjamini-Hochberg). The input into the FDR algorithm was not all the p-values in the experiment, instead we applied FDR correction for the array of p-values associated with an electrode for a participant's energy.

Overfitting was not a concern for the permutation analysis. The entire dataset was used for both the permutations and true data. Therefore, if the true model overfit the data, the same possibility of overfitting would exist for the permuted data. However, the data were low-dimensional with only 1 feature, the score, for approximately 80 data points (trials). Therefore, the model had poor fits on the permuted data. In high-dimensional, sparse data, it is far easier for a model to overfit.

Our final analysis consisted of Random Walk Metropolis-Hasting (MH) sampling the beta coefficients used for the probit regression. At each ETE point, the score was calculated and a modified version of the probit classifier was employed.

We then ran the Metropolis-Hasting (MH) algorithm for 500 iterations, and discarded the first 200 as burn in. The starting value of the coefficient for the first iteration was 0. At each iteration of the MH algorithm, a candidate was sampled from a normal distribution with the previous iteration's coefficient as the mean: $\beta^{candidate} \sim N(\beta^{previous}, .1)$. The binomial log-likelihood associated with $\beta^{candidate}$ was compared with the binomial log-likelihood associated with $\beta^{previous}$. If the difference was larger than the log of a random number between 0 and 1, then the current iteration's β was set to $\beta^{candidate}$, otherwise it was set to $\beta^{previous}$.

The MH algorithm resulted in a length 300 distribution for each ETE point. From this distribution we calculated the mean coefficient, and we calculated the number of iterations for which the coefficient was greater than or less than 0.

The p-values from the permutation analyses and the coefficient sampling further corroborate which of the ETE combinations distinguish between correct and incorrect responses. For each participant we identified statistically significant ETE combinations with the highest accuracy in classifying responses as correct/incorrect. We chose ETE combinations that showed the highest accuracy for a duration of at least 10 ms to increase the validity of the identified ETE combinations as compared to the smallest possible

window of 5 ms. By prolonging the time window to 10 ms we sought to decrease the likelihood of randomly significant ETE combinations, and also to increase the possible clinical translation as we assume that longer windows will be easier to target therapeutically than shorter windows. Last, we checked if the p-values of the identified ETE combinations survived FDR correction.

3. Results

3.1. Naming task performance

After removing segments with moderate or excessive motion or movement artifact (bad blocks) from the EEG signal, at least 79 of 80 total naming trials remained for each participant (Table 2). The five participants varied in their naming task performance. Participant 2 produced the most correct responses with 70% of all stimuli being named correctly, and participant 3 produced the least correct responses with 30% of all stimuli being named correctly (Table 2). The type of incorrect response also varied within and between participants. Importantly, there were overall only few self-corrections (Participants 1 and 2: one, all other participants: no self-corrections) or no responses (Participant 1: one, participant 2: zero, participant 3: eight, participant 4: four, participant 5: five no responses); thus, the vast majority of incorrect responses consisted of articulated responses that were not being corrected. For all participants, response latencies were on average longer for incorrect than for correct responses. Across participants the average naming latencies ranged from 1929 ms for correct responses (participant 5) to 6005 ms for incorrect responses (participant 2) (Table 2).

3.2. Prediction of correct/incorrect responses with machine learning

The ETE combinations with the highest accuracy to predict correct compared to incorrect responses all varied between

participants and ranged from 63% (participant 5) to 80% (participant 4). Table 3 shows the best ETE combination for each participant and model performance values (e.g. sensitivity, specificity). Fig. 4 shows the location of the electrodes that were part of the best ETE combinations. These were in the superior or middle parietal lobe for participants 1, 4 and 5 (P1, CP1 and P5, respectively), and in the superior to middle frontal lobe for participants 2 and 3 (FC3 and F3, respectively). The time ranges for the best ETE combinations were within 1000 ms after stimulus presentation for all participants. ETE combinations including beta energy bands had the best predictive accuracy for participants 1 and 2, while alpha energy bands had the best predictive accuracy for participants 3, 4 and 5. The ETE combinations and their corresponding accuracies for alpha and beta energy bands are shown in the heat maps of Fig. 5. Fig. 6 shows raw p-values and FDR corrected p-values for the energy band that included the electrode/time-range/energy (ETE) combination(s) with the highest accuracy. For all but one participant the p-values for the prediction accuracy survived FDR correction. The prediction accuracy from participant 5 was not better than chance when the true accuracy score was compared with the accuracy permutation distribution. However, when the true AUROC score was compared with the AUROC permutation distribution, the prediction accuracy from participant 5 (as well as from all other participants) was higher than chance. For participant 2 only one 5 ms window (at 945 ms after stimulus presentation) survived FDR correction, compared to participants 1, 3, and 4, for whom time windows of at least 10 ms survived FDR correction.

4. Discussion

The goal of this project was to investigate whether pre-articulatory neural activity can be used systematically to predict correct vs. incorrect naming responses in individuals with post-stroke aphasia. Using machine learning, we investigated the prediction accuracy of hEEG recordings of five participants who

Table 2
Naming responses for each participant (depending on the number of naming trials that had to be excluded based on artifacts in the EEG signal, the total number of responses may not sum up to 80 concrete stimuli).

Response type		Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
Correct	Number (%)	30 (38)	55 (70)	24 (30)	49 (61)	38 (48)
	Response latency in ms, average (range)	2320 (745–6847)	5583 (2271–7212)	1975 (955–7150)	4215 (2,952–8015)	1929 (830–6700)
Incorrect	Number (%)	49 (62)	24 (30)	55 (70)	31 (39)	42 (52)
	Response latency in ms, average (range)	2542 (881–5777)	6005 (4925–8728)	2613 (1027–6358)	5051 (1786–7857)	2658 (1196–7255)
Total number of included naming trials (max: 80)		79	79	79	80	80

Table 3
Electrode/time-range/energy (ETE) combinations with the highest accuracy to predict correct compared to incorrect responses for each participant. Only ETEs are listed with the highest accuracy for ≥ 10 ms long windows. ≥ 10 ms long windows were chosen to increase the validity of significant ETE combinations as compared to shorter 5 ms long windows. All listed ETE combinations remained significant after false discovery rate (FDR) correction, except for the ETE combination of participant 5. AUROC = area under the receiver operating characteristics, TPR = true positive rate, TNR = true negative rate, PPV = positive predictive value, NPV = negative predictive value.

	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
Prediction accuracy	0.71–0.72	0.68–0.69	0.71–0.75	0.71–0.80	0.63–0.66
Electrode	P1	FC3	F3	CP1	P5
Time range (milliseconds)	830–865	930–955	490–570	380–685	850–880
Energy band	beta	beta	alpha	alpha	alpha
AUROC	0.75 (0.73–0.77)	0.62 (0.62–0.64)	0.74 (0.72–0.75)	0.72 (0.66–0.75)	0.69 (0.68–0.71)
TPR (Sensitivity)	0.68 (0.63–0.70)	0.72 (0.71–0.73)	0.69 (0.67–0.71)	0.78 (0.71–0.88)	0.63 (0.63–0.63)
TNR (Specificity)	0.73 (0.69–0.80)	0.63 (0.63–0.63)	0.73 (0.71–0.76)	0.68 (0.65–0.71)	0.67 (0.62–0.69)
PPV	0.62 (0.57–0.66)	0.82 (0.81–0.82)	0.53 (0.50–0.57)	0.80 (0.77–0.82)	0.63 (0.60–0.65)
NPV	0.79 (0.77–0.80)	0.49 (0.48–0.50)	0.85 (0.83–0.86)	0.67 (0.61–0.78)	0.67 (0.65–0.67)

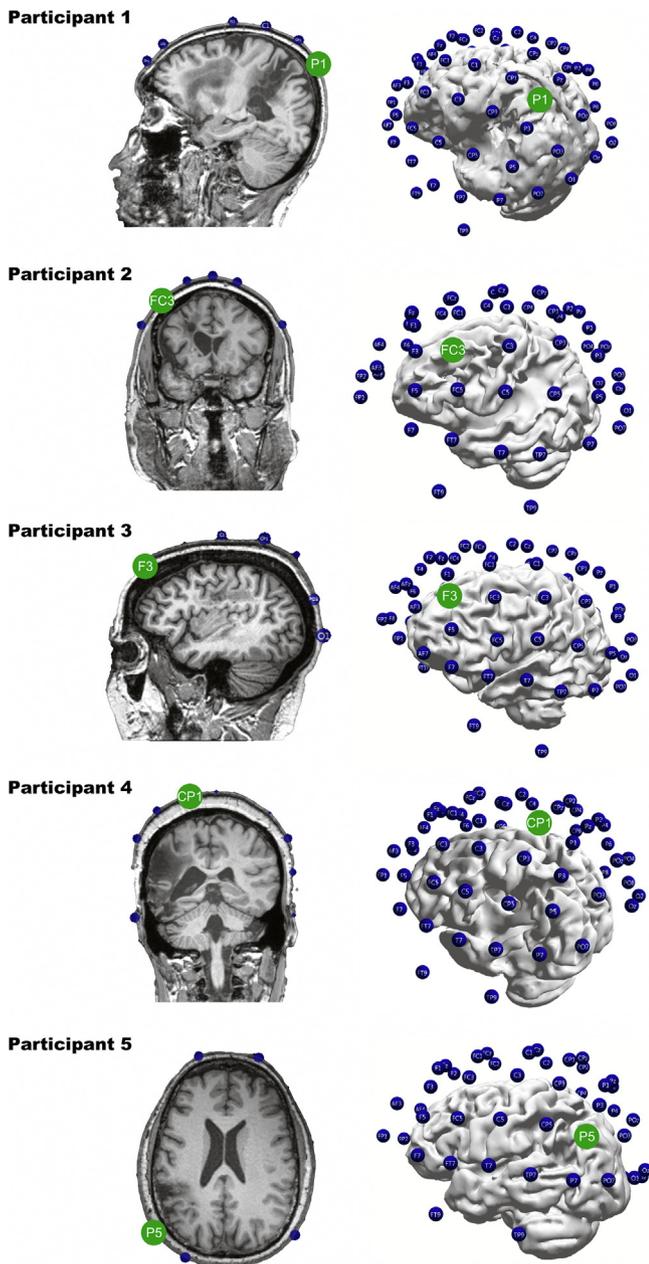


Fig. 4. Location of the electrode with the best electrode/time-range/energy (ETE) combination (highlighted in green). First column shows the participant's MRI (sagittal view for participants 1 and 3, coronal view for participants 2 and 4, axial view for participant 5); second column shows the left view of the reconstructed participant-specific 3D head model (viewing angle was chosen based on best visibility of the electrode and stroke lesion).

completed a naming task. Our results indicate that in every individual there are ETE combinations associated with correct vs incorrect responses, however, the success of predicting correct/incorrect responses based on this pattern of pre-articulatory neural activity varies between individuals. While we found ETE combinations with accuracies of 0.7 to 0.8 for three of five individuals, the ETE combination of one individual did not reach 0.7, and the ETE combination of another individual was not better than chance. However, predictions were better than chance for all individuals based on AUROC values. In contrast to accuracy values, AUROC values account for the sensitivity and specificity of predictions and are thus, sometimes discussed as a more reliable measure.

The predictive accuracy for correct vs. incorrect responses did not seem to depend on the occurrence of apraxia of speech or

dysarthria. One of the two participants (participant 2) had moderate-severe evidence of both, apraxia of speech and dysarthria, whereas another participant (participant 5) had no evidence of either apraxia of speech or dysarthria. The two participants had similar accuracy and area under the receiver operating characteristics (AUROC) values.

It is worth noting that the differences in PPV or NPV between participants is influenced by prevalence as well as model performance. For example, the positive predictive value (PPV) in participant 2 was high (0.82), but the negative predictive value (NPV) low (0.49). This was the opposite case for participant 3 (PPV = 0.53, NPV = 0.73). Participant 2 produced 70% correct naming responses, and participant 3 produced 70% incorrect naming responses.

An ETE combination that had the highest predictive power for one participant did not repeat in another participant. The five participants in our study varied in terms of which electrode, time range, and energy band had the highest prediction accuracy. Thus, ETE combinations with the highest predictive accuracy are not generalizable across participants and need to be determined individually. This is in line with previous research suggesting that a baseline EEG signature of correct and incorrect naming responses needs to be acquired from each participant individually before EEG signals could be used to tailor treatment (Riley et al., 2017). For example, the first step of a treatment process could be to apply the machine learning model presented in this study to identify ETE combinations with the highest predictive accuracy for correct vs. incorrect responses. In a second step, treatment approaches, such as neuromodulation techniques, could be tailored to the individual participant by using the identified ETE combinations.

As to be expected, for all participants the electrodes of the best ETE combinations were located outside the lesion. Based on our study hypothesis, we anticipated the location of the electrodes will be around the stroke lesion – in perilesional areas of the left hemisphere. This was the case for most of the ETE combinations. For example, electrode P1 showed the highest predictive accuracy in participant 1, electrode FC3 in participant 2, and electrode P5 in participant 5. These electrodes were located directly adjacent to the lesion (Fig. 4). The increasing importance of perilesional areas after a stroke has been suggested in previous research showing that aphasia recovery is associated with increased perilesional brain activity (Meinzer et al., 2004; Lidzba et al., 2012). However, electrode F3 showed the highest predictive accuracy in participant 3 and electrode CP1 in participant 4. Both electrodes were not located adjacent to the lesion. Thus, our results suggest that perilesional brain areas may be crucial for correct vs incorrect naming responses, but this observation is not generalizable across patients. A possible explanation for differences in the proximity of the electrodes to the lesion is the integrity of the remaining, non-lesioned, brain tissue in the left hemisphere (Fridriksson et al., 2010; Bonilha et al., 2016). If a stroke spares crucial language areas (no matter if they are close or distant from the lesion), it seems likely that these areas maintain or even increase their leading roles in language performance after the stroke. In our study it remains unknown whether the areas with the best ETE combinations had a crucial role for naming performance already before the stroke, or if these areas at least partially took over the role of brain areas that had been damaged by the stroke lesion. Future longitudinal studies could shed light on these observations by assessing participants' language performance and brain activity at multiple time points to understand the relationship between language recovery, neurophysiological changes and their impact on the accuracy of prediction models for naming errors.

ETE combinations with the highest predictive accuracy were mostly in time windows later than 500 ms after stimulus presentation. Research on healthy and aphasic individuals identified initial neural activation (~0–200 ms) during visual processing, followed

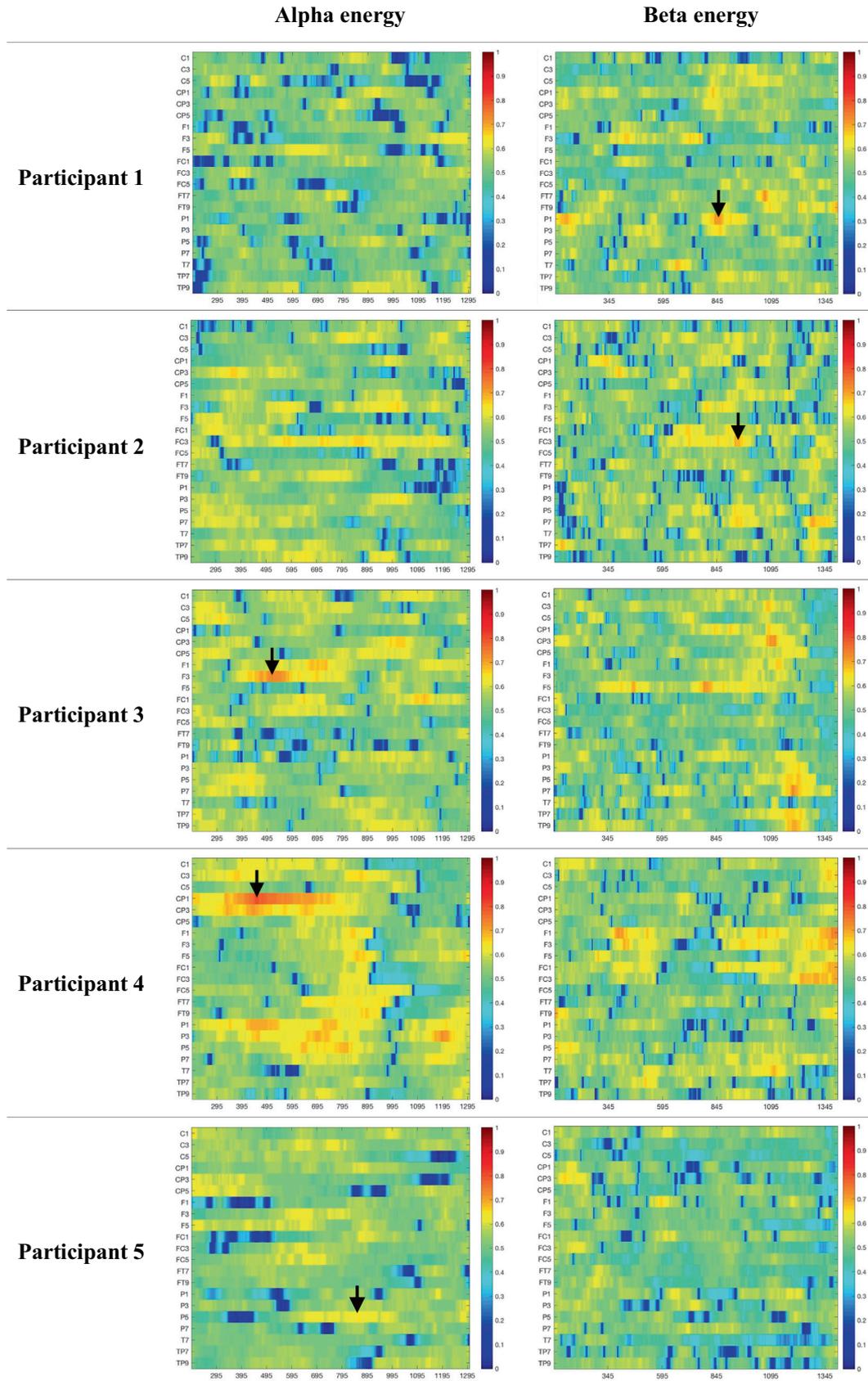


Fig. 5. Heat maps of electrode/time-range/energy (ETE) combinations and their corresponding accuracies in predicting correct/incorrect responses for alpha and beta energy bands. The x-axes show the time in milliseconds after the stimulus presentation with time “0” referring to the onset of stimulus presentation. The y-axes refer to the included EEG electrodes from the left hemisphere (in alphabetical order). Different colors represent the accuracy value, with warmer colors indicating higher accuracies. Black arrows indicate the ETE combination with the highest accuracy for each participant (see Table 3).

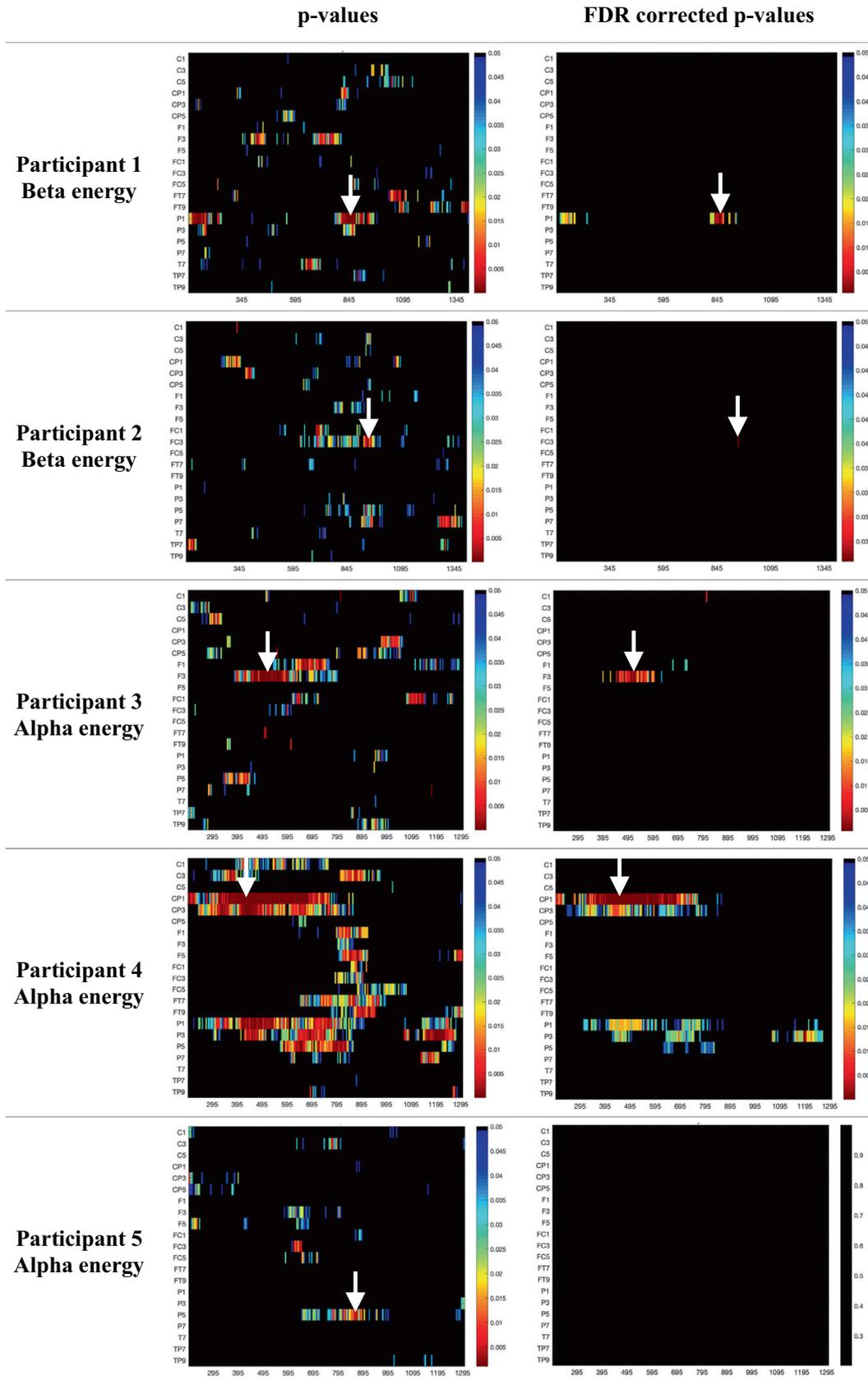


Fig. 6. Heat maps for each participant showing p-values for the energy band that includes the electrode/time-range/energy (ETE) combination(s) with the highest accuracy in predicting correct/incorrect responses (see Table 3). The first column of heat maps shows uncorrected p-values, the second column shows p-values corrected for false discovery rate (FDR). The x-axis of each heat map shows the time in milliseconds after the stimulus presentation with time “0” referring to the onset of stimulus presentation. The y-axis of each heat map refers to the included EEG electrodes from the left hemisphere (in alphabetical order). Different colors represent p-values < 0.05, with warmer colors indicating lower p-values. White arrows indicate the ETE combination with the highest accuracy for each participant (see Table 3).

by semantic (~100 ms–300 ms) and phonological processing (~250 ms– 00 ms) (Vihla et al., 2006; Indefrey, 2011; Laganaro et al., 2013; Singh et al., 2018). Amongst our cohort, this finding may suggest that most of the differentiation in neural activation between correct and incorrect responses occurs in later word processing stages. However, response times varied greatly between and within participants, ranging from less than 1 second to more than 8 seconds. This indicates that the processing stages for language production during naming also varied between and within participants. For example, the best ETE combination for participants 1 and 2 were in similar time windows (830–865 ms and 930–940 ms, respectively), but the average response time for correct and incorrect naming responses showed a large difference of approximately 3 seconds. Thus, it remains speculative at what processing stage the ETE windows occurred in each participant and trial. Future studies should include a thorough assessment of participants' time-dependent language processing to allow a better understanding of the underlying neurolinguistic mechanisms of the ETE combinations.

Our study has demonstrated that pre-articulatory activity is a neurophysiological pattern that can be used to predict naming responses in individuals with aphasia. Future research is needed to further improve our methods to boost prediction accuracy and extend the applicability to more detailed error classifications (e.g., semantic errors, phonological errors, no response). We hope to contribute in the future methodological improvements by analyzing data from a larger group of participants. The significance of this line of research is to make use of brain activity before an individual utters a response. If we can understand what occurs in the brain when an individual with aphasia names an object correctly before the response is even produced, then perhaps this information can be incorporated into a treatment regimen for patients; for example, to modulate brain activity with closed-loop neuromodulation techniques to enhance the likelihood of correct responses.

4.1. Limitations

We investigated a small sample of individuals with aphasia, which does not allow for systematic between-subject analyses; such analyses could include the effect of the stroke location, time since stroke, and aphasia type on differences in the prediction models and accuracies. Because our study was exploratory, we only used a limited number of stimuli (80). Therefore, we simplified the error classifications to correct vs incorrect. A more precise error classification may provide important information on the EEG signal associated with specific errors (e.g. semantic errors, phonemic errors) and response patterns (e.g. no response, self-correction), because the error types might differ in their brain activity. This should be the target of future studies. Moreover, we did not analyze gamma energy bands and analyzed alpha and beta energy only. Previous research using electrocorticography (ECoG) has identified gamma energy bands to be significantly associated with naming responses (Sinai et al., 2005; Tanji et al., 2005). Nevertheless, we had to exclude gamma energy, because when using scalp EEG (as in our study), gamma energy can easily be contaminated by possible (muscle) artifacts (Muthukumaraswamy, 2013). Finally, performance in confrontational (object) naming tasks may not be entirely transferable to assessing everyday naming or communication abilities despite wide use in clinical assessments, treatments, and research.

Our performance results are better than chance and an excellent start, but the results were still modest. Accuracies were mostly around 70% and for two individuals (participant 2 and participant 5) no ETE combinations of at least 10 ms duration survived FDR correction. We anticipate that including spatiotemporal information

can improve the predictive capability of the model. For example, the accuracy heatmaps show consistency in nearby time points. Future steps should include methods that incorporate spatiotemporal information, such as convolution.

5. Conclusions

Our findings indicate that it is possible to predict correct and incorrect naming responses based on pre-articulatory neural activity. ETE combinations with the highest prediction accuracy vary across participants and need to be determined individually. If future research can improve prediction accuracy further, we believe that this line of research has the potential to guide the development of new treatment approaches that take neural activity into consideration.

Acknowledgements

This study was supported by research grants from the National Institutes of Health/National Institute on Deafness and Other Communication Disorders (NIDCD) [grant numbers DC014021, DC011739, DC014664, DC015831-02], and from the American Heart Association [grant number SFDNR26030003].

Declaration of Competing Interest

None.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.clinph.2019.08.011>.

References

- Baldo JV, Arevalo A, Patterson JP, Dronkers NF. Grey and white matter correlates of picture naming: evidence from a voxel-based lesion analysis of the Boston Naming Test. *Cortex* 2013;49:658–67.
- Boersma P, Weenink D. Praat: doing phonetics by computer. 6.0.37 ed2018.
- Bonilha L, Gleichgerrcht E, Nesland T, Rorden C, Fridriksson J. Success of anomia treatment in aphasia is associated with preserved architecture of global and left temporal lobe structural networks. *Neurorehabil Neural Repair* 2016;30:266–79.
- DeLeon J, Gottesman RF, Kleinman JT, Newhart M, Davis C, Heidler-Gary J, et al. Neural regions essential for distinct cognitive processes underlying picture naming. *Brain* 2007;130:1408–22.
- Ezzyat Y, Wanda PA, Levy DF, Kadel A, Aka A, Pedisich I, et al. Closed-loop stimulation of temporal cortex rescues functional networks and improves memory. *Nat Commun* 2018;9:365.
- Freed DB, Marshall RC, Chuhlantseff EA. Picture naming variability: a methodological consideration of inconsistent naming responses in fluent and nonfluent aphasia. *Clin Aphasiol* 1996;24:193–206.
- Fridriksson J, Bonilha L, Baker JM, Moser D, Rorden C. Activity in preserved left hemisphere regions predicts anomia severity in aphasia. *Cereb Cortex* 2010;20:1013–9.
- Fridriksson J, Richardson JD, Fillmore P, Cai B. Left hemisphere plasticity and aphasia recovery. *Neuroimage* 2012;60:854–63.
- Gleichgerrcht E, Fridriksson J, Bonilha L. Neuroanatomical foundations of naming impairments across different neurologic conditions. *Neurology* 2015;85:284–92.
- Howard D, Patterson K, Franklin S, Morton J, Orchard-Lisle V. Variability and consistency in picture naming by aphasic patients. *Adv Neurol* 1984;42:263–76.
- Humphreys GW, Price CJ, Riddoch MJ. From objects to names: a cognitive neuroscience approach. *Psychol Res* 1999;62:118–30.
- Indefrey P. The spatial and temporal signatures of word production components: A critical update. *Front Psychol* 2011;2:255.
- Kertesz A. The western aphasia battery - revised. New York: Grune & Stratton; 2007.
- Laganaro M, Morand S, Michel CM, Spinelli L, Schnider A. ERP correlates of word production before and after stroke in an aphasic patient. *J Cogn Neurosci* 2011;23:374–81.
- Laganaro M, Morand S, Schnider A. Time course of evoked-potential changes in different forms of anomia in aphasia. *J Cogn Neurosci* 2009;21:1499–510.

- Laganaro M, Python G, Toepel U. Dynamics of phonological-phonetic encoding in word production: evidence from diverging ERPs between stroke patients and controls. *Brain Lang* 2013;126:123–32.
- Lidzba K, Staudt M, Zieske F, Schwilling E, Ackermann H. Prestroke/poststroke fMRI in aphasia: Perilesional hemodynamic activation and language recovery. *Neurology* 2012;78:289.
- Meinzer M, Elbert T, Wienbruch C, Djundja D, Barthel G, Rockstroh B. Intensive language training enhances brain plasticity in chronic aphasia. *BMC Biol* 2004;2:20.
- Muthukumaraswamy SD. High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Front Hum Neurosci* 2013;7:138.
- Nozari N, Kittredge AK, Dell GS, Schwartz MF. Naming and repetition in aphasia: Steps, routes, and frequency effects. *J Mem Lang* 2010;63:541–59.
- Riley EA, McFarland DJ. EEG error prediction as a solution for combining the advantages of retrieval practice and errorless learning. *Front Hum Neurosci* 2017;11.
- Seo NJ, Lakshminarayanan K, Lauer AW, Ramakrishnan V, Schmit BD, Hanlon CA, et al. Use of imperceptible wrist vibration to modulate sensorimotor cortical activity. *Exp Brain Res* 2019.
- Sinai A, Bowers CW, Crainiceanu CM, Boatman D, Gordon B, Lesser RP, et al. Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming. *Brain* 2005;128:1556–70.
- Singh T, Phillip L, Behroozmand R, Gleichgerrcht E, Piai V, Fridriksson J, et al. Pre-articulatory electrical activity associated with correct naming in individuals with aphasia. *Brain Lang* 2018;177–178:1–6.
- Strand EA, Duffy JR, Clark HM, Josephs K. The Apraxia of Speech Rating Scale: a tool for diagnosis and description of apraxia of speech. *J Commun Disord* 2014;51:43–50.
- Tanji K, Suzuki K, Delorme A, Shamoto H, Nakasato N. High-frequency gamma-band activity in the basal temporal cortex during picture-naming and lexical-decision tasks. *J Neurosci* 2005;25:3287–93.
- Vihla M, Laine M, Salmelin R. Cortical dynamics of visual/semantic vs. phonological analysis in picture confrontation. *Neuroimage* 2006;33:732–8.