Detecting post-stroke aphasia using EEG-based neural envelope tracking of natural speech

Pieter De Clercq[1], Jill Kries[1], Ramtin Mehraram[1], Jonas Vanthornhout[1], Tom Francart[1] and Maaike Vandermosten[1]

[1]: Experimental Oto-Rhino-Laryngology, Department of Neurosciences, Leuven Brain Institute, KU Leuven, Belgium

Version:

March 2023

Corresponding author:

Pieter De Clercq (pieter.declercq@kuleuven.be)

Abstract

After a stroke, approximately one-third of patients suffer from aphasia, a language disorder that impairs communication ability. The standard behavioral tests used to diagnose aphasia are time-consuming, require subjective interpretation, and have low ecological validity. As a consequence, comorbid cognitive problems present in individuals with aphasia (IWA) can bias test results, generating a discrepancy between test outcomes and everyday-life language abilities. Neural tracking of the speech envelope is a promising tool for investigating brain responses to natural speech. The envelope of speech is crucial for speech understanding, encompassing cues for detecting and segmenting linguistic units, e.g., phrases, words and phonemes. In this study, we aimed to test the potential of the neural envelope tracking technique for detecting language impairments in IWA.

We recorded EEG from 27 IWA in the chronic phase after stroke and 22 healthy controls while they listened to a 25-minute story. We quantified neural envelope tracking in a broadband frequency range as well as in the delta, theta, alpha, beta, and gamma frequency bands using mutual information analysis. Besides group differences in neural tracking measures, we also tested its suitability for detecting aphasia at the individual level using a Support Vector Machine (SVM) classifier. We further investigated the required recording length for the SVM to detect aphasia and to obtain reliable outcomes.

IWA displayed decreased neural envelope tracking compared to healthy controls in the broad, delta, theta, and gamma band, which is in line with the assumed role of these bands in auditory and linguistic processing of speech. Neural tracking in these frequency bands effectively captured aphasia at the individual level, with an SVM accuracy of 84% and an area under the curve of 88%. Moreover, we demonstrated that high-accuracy detection of aphasia can be achieved in a time-efficient (5 minutes) and highly reliable manner (split-half reliability correlations between R=0.62 and R=0.96 across frequency bands).

Our study shows that neural envelope tracking of natural speech is an effective biomarker for language impairments in post-stroke aphasia. We demonstrated its potential as a diagnostic tool with high reliability, individual-level detection of aphasia, and time-efficient assessment. This work represents a significant step towards more automatic, objective, and ecologically valid assessments of language impairments in aphasia.

Keywords: Aphasia, natural speech processing, neural envelope tracking, diagnostics

1 Introduction

Aphasia is an acquired language disorder impairing communication ability and is principally caused by a stroke in the language-dominant left hemisphere (Papathanasiou and Coppens, 2017). The current practice is to diagnose aphasia by means of behavioral language tests. However, these tests suffer from influences of co-morbid motor and cognitive problems (Rohde et al., 2018) and of a low ecological validity (Devanga et al., 2021; Wallace and Kimelman, 2013). Novel analysis techniques for EEG-data, i.e., neural tracking of speech (e.g., see Lalor et al. (2009); Brodbeck et al. (2022); Crosse et al. (2021)), allow measuring brain responses while participants listen to natural speech, providing an ecologically valid way to measure speech processing. In this study, we test the potential of these novel EEG analyses for detecting language impairments in aphasia with high accuracy and in a time-efficient way.

The current standard for diagnosing aphasia is based on performance on behavioral language tests, such as the Western Aphasia Battery (Kertesz, 1982), the Token test (de Renzi and Ferrai, 1978) or a picture-naming test (Van Ewijk et al., 2020). Yet, these tests have several disadvantages. First, behavioral testing is time-consuming, requiring active cooperation of the patient and scoring by the clinician. Second, behavioral assessment can lead to inaccurate initial diagnoses due to concomitant motor, memory, attention and executive impairments (Rohde et al., 2018), reportedly affecting over 80% of individuals with aphasia (IWA) (El Hachioui et al., 2014). Finally, language tests consist of rather artificial tasks in which sounds, phonemes, words or short sentences are presented in isolation. This contrasts with natural speech processing where language components interact, and higher-level context integration takes place (Hamilton and Huth, 2018; Kandylaki and Bornkessel-Schlesewsky, 2019). Consequently, there is a discrepancy between clinical assessment and a patient's natural speech abilities in everyday life (Lesser and Algar, 1995; Kim et al., 2022; Stark et al., 2021; Wallace and Kimelman, 2013).

EEG-based event-related potential (ERP) studies have been conducted to address limitations of behavioral testing. Studies have shown that IWA exhibit altered ERP components such as the P1, N1, P2, N2, P300, and N400 in response to language stimuli (Aerts et al., 2015; Becker and Reinvang, 2007; Ilvonen et al., 2001; Ofek et al., 2013; Pulvermüller et al., 2004; Robson et al., 2017). Together with the potential for automatic assessment that requires less active participation from the patient, these alterations suggest that ERPs may have diagnostic value in aphasia (Cocquyt et al., 2020). Nonetheless, ERP paradigms involve artificial language stimuli presented repeatedly to the participant, which questions the ecological validity of the obtained outcomes (Le et al., 2018). Furthermore, previous ERP studies in aphasia have not reported on the reliability of ERPs, the minimal required recording length and the sensitivity to capture language impairments at the individual level.

Recent studies have investigated the EEG response to natural, running speech, which could open new perspectives to studying natural speech processing in IWA. When listening to speech, the brain tracks

the temporal envelope, which contains essential cues for speech understanding. The envelope consists of the slow-varying temporal modulations in the speech signal and encompasses cues for detecting and identifying lexical units (i.e., phonemes, syllables, words and phrases) and prosody (Peelle and Davis, 2012). In fact, previous research showed that listeners can understand speech based on the low-frequency temporal envelope only (Shannon et al., 1995). Neural envelope tracking can be measured by applying encoding and decoding models on the stimulus and the recorded EEG. The level of tracking is reflected in the extent to which the models can either predict the neural signals or decode the envelope. In a linear (Crosse et al., 2021) or a mutual information-based (De Clercq et al., 2023) model, the neural tracking outcomes can be visualized over time and space (i.e., EEG channels), obtaining response properties similar to traditional ERP components (Brodbeck et al., 2022). The neural tracking technique is rapidly evolving and has led to crucial new insights as to how natural speech is processed in the brain. Neural envelope tracking is strongly related to speech understanding (Ding and Simon, 2013; Etard and Reichenbach, 2019; Kaufeld et al., 2020), and can be used to objectively quantify speech intelligibility (Vanthornhout et al., 2018; Gillis et al., 2022).

Prior research assigned the low-frequency temporal envelope primarily to speech understanding. The low-frequency envelope, i.e., delta (0.5–4 Hz) and theta (4-8 Hz) band, encompass cues for detecting and segmenting lexical units. The theta band tracks syllables and lower-level acoustic processing of speech (Etard and Reichenbach, 2019), while the delta band signal is associated with processing speech prosody and segmenting higher-level linguistic structures such as words and phrases (Ding et al., 2016; Giraud and Poeppel, 2012; Kaufeld et al., 2020). In addition to the delta and theta band, reflecting synthesis of higher-level auditory and linguistic structures, the alpha and beta bands are involved in attention and auditory-motor coupling (Wöstmann et al., 2017; Fujioka et al., 2015), while the gamma band is involved in encoding phonetic features (Giraud and Poeppel, 2012; Gross et al., 2013; Hyafil et al., 2015). In conclusion, specific frequency bands are believed to reflect different stages of speech processing.

Neural envelope tracking of natural speech has been investigated in several clinical populations with language impairments. For individuals with primary progressive aphasia, a language disorder caused by a neurodegenerative disease, Dial et al. (2021) reported increased neural tracking in the theta band but no group differences in the delta band. The researchers argued that enhanced theta band tracking in individuals with primary progressive aphasia might reflect a compensation mechanism through increased reliance on acoustic cues. For individuals with dyslexia, a disorder characterized by phonological processing difficulties, decreased tracking in delta, theta and beta/gamma (phoneme- and phonetic-level) band have been reported (Di Liberto and Lalor, 2017; Lizarazu et al., 2021; Mandke et al., 2022). In conclusion, these studies have shown the potential for neural tracking to capture language impairments in clinical cohorts.

The present study investigated whether we can differentiate IWA in the chronic phase after stroke

(i.e., $\geq$ 6 months post-stroke) from neurologically healthy, age-matched controls using EEG-based neural envelope tracking. Specifically, we used mutual information analyses to quantify neural envelope tracking, which captures linear and nonlinear effects and outperforms linear models (De Clercq et al., 2023). We described both groups' responses to the speech envelope temporally and spatially at broadband frequency range. We further investigated neural tracking in specific frequency bands ranging from delta to gamma band, as different frequency bands are involved in different (sub-)lexical processes (Etard and Reichenbach, 2019; Ding et al., 2016; Giraud and Poeppel, 2012; Keitel et al., 2018; Peelle and Davis, 2012).

Secondly, we assessed the suitability of the neural tracking technique as a biomarker to capture language processing difficulties. To this end, we used a Support Vector Machine (SVM) to classify participants as healthy or aphasic using MI measures in different frequency bands as input to the model. Finally, we investigated how much data the neural tracking technique requires for good classification and reliable outcomes.

2 Materials and methods

2.1 Participants

Our sample comprised 27 IWA (seven female participants, 73 $\pm$11 y/o) in the chronic phase ($\geq$ 6 months) after stroke and 22 neurologically healthy controls (seven female participants, 72 $\pm$7 y/o). There was no significant age difference between groups (unpaired Wilcoxon rank sum test: W=343.5, p=0.36). IWA were recruited at the stroke unit of the University Hospital Leuven and via speech-language pathologists. Healthy controls were recruited, making sure they matched the age of IWA at the group level. The inclusion criteria for IWA were: (1) a left-hemispheric or bilateral stroke, (2) a diagnosis of aphasia in the acute stage after stroke using behavioral language tests and (3) no formal diagnosis of a psychiatric or neurodegenerative disorder. For more information regarding demographics, recruitment strategy and diagnosis in the acute stage after stroke, we refer to Kries et al. (2022). The study was approved by the ethical committee UZ/KU Leuven (S60007), and all participants gave written consent before participation. Research was conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements.

Participants completed standardized clinical tests for aphasia at the time of participation as described in detail in Kries et al. (2022). IWA scored significantly lower on the 'Nederlandse Benoemtest', i.e., Dutch Naming Test (Van Ewijk et al., 2020), and the ScreeLing test (El Hachioui et al., 2017; Visch-Brink et al., 2010) compared to healthy controls (W=57.5, p<0.001; W=101, p<0.001, respectively). Although seven IWA did not score below the cut-off threshold for aphasia on either of these tasks, they were still attending speech-language therapy sessions at the time of participation and had extended documentation

of language deficits in the acute stage after stroke (Kries et al., 2022).

2.2 EEG experiment

The EEG measurements took place in a soundproof, electromagnetically shielded booth using a 64-channel BioSemi ActiveTwo system (Amsterdam, the Netherlands) at a sampling frequency of 8,192 Hz. Participants were instructed to listen to a 25-minute-long story, *De Wilde Zwanen*, written by Christian Andersen and narrated by a female Flemish-native speaker, presented in silence while their EEG was recorded. The story was cut into five parts with an average duration of 4.84 minutes. After each story part, participants answered content questions about the preceding part, introduced to make the participant follow the content attentively. Participants had a short break after each story part and answered content questions about the preceding part. The protocol introduced these questions to make participants follow the story attentively. The story was presented bilaterally through ER-3A insert earphones (Etymotic Research Inc, IL, USA) using the software platform APEX (Francart et al., 2008).

We determined a subject-dependent intensity level at which the story was presented based on the thresholds of the pitch tone audiometry (PTA). We defined hearing thresholds for octave frequencies between .25 and 4 kHz. For normal hearing participants, the story was presented at 60 dBA. For hearing impaired participants, defined as participants that have a hearing threshold >25 dB hearing loss on frequencies below 4 kHz, the volume was augmented with half of the pure tone average of the individual thresholds at .25, .5 and 1 kHz for both ears individually. This procedure was adapted from Jansen et al. (2012). To check whether age-related hearing loss differed between both groups, we calculated the Fletcher index, i.e., average of PTA thresholds at .5, 1 and 2 kHz. Hearing levels did not differ between groups (Fletcher index averaged across the right and left ear: W=326.5, p=0.56).

2.3 Signal processing

Envelope extraction

We used a gammatone filter bank (Søndergaard et al., 2012) to extract the envelope. We used 28 channels spaced by one equivalent rectangular bandwidth and center frequencies from 50 Hz until 5000 Hz. The envelopes were extracted from each sub-band by taking the absolute value of each sample and raising it to the power of 0.6. The resulting 28 sub-band envelopes were averaged to obtain a single envelope. Next, the envelope was downsampled to 512 Hz to decrease processing time. The envelope was then filtered in frequency ranges of interest. These include delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), low-gamma (30-49 Hz) and a broad (0.5-49 Hz, including all individual frequency ranges) band. We used high- and lowpass filters, with a transition band of 10% below the highpass and 10% above the lowpass frequency. A Least Squares filter of order 2000 was used, and we compensated for the group delay. After filtering, the envelope was normalized and further downsampled to 128 Hz.

EEG data processing

EEG data were pre-processed using the Automagic toolbox (Pedroni et al., 2019) and custom Matlab scripts (The MathWorks Inc., Natick, MA, USA, 2021). The EEG signals were first downsampled to 512 Hz to decrease processing time. Artifacts were removed using the artifact subspace reconstruction method (Mullen et al., 2015). Next, an independent component analysis was applied to the data, and components classified as "brain" or "other" (i.e., mixed components), using the EEGLAB plugin ICLabel (Pion-Tonachini et al., 2019), with a probability higher than 50% were preserved (average number of removed components: 26 ±7). The neural signals were projected back to the channel space, where the signals were average referenced. Subsequently, we filtered the EEG data in the same frequency bands using the same Least Squares filter as in the envelope extraction method. Next, normalization and further downsampling to 128 Hz were applied.

2.4 Neural envelope tracking

We investigated neural envelope tracking using the Gaussian copula MI analysis (Ince et al., 2017). In the Gaussian copula approach, all variables (the envelope and EEG channels) are first ranked on a scale from 0 to 1, obtaining the cumulative density function (CDF). By computing the inverse standard normal CDF, the data distributions of all variables are transformed to perfect standard Gaussians. Subsequently, the parametric Gaussian MI estimate can be applied to the data provided by:

$$I(X;Y) = \frac{1}{2ln2}\left[\frac{|\sum_X||\sum_Y|}{|\sum_{XY}|}\right] \quad (1)$$

where I(X;Y) equals the MI between X and Y (here, the EEG and the envelope), expressed in bits. $|\sum_X|$ and $|\sum_Y|$ are the determinants of the covariance matrices of X and Y, and $|\sum_{XY}|$ is the determinant of the covariance matrix for the joint variable. To obtain temporal information on MI, we shifted the EEG as a function of the envelope over time (using an integration window -200 to 500 ms) and applied Eq. (1) at each sample. The result forms the temporal mutual information function (TMIF) and reflects how the brain processes speech over time (De Clercq et al., 2023; Zan et al., 2020). For an in-depth explanation of the Gaussian copula MI method, we refer to Ince et al. (2017). For a more practical explanation of the TMIF in the context of neural envelope tracking, see De Clercq et al. (2023).

We calculated the single-channel TMIF and the multivariate TMIF, analog to a (linear) encoding and decoding model, respectively. The single-channel TMIF calculates the TMIF for each channel individually, providing both temporal (i.e., peak latency and peak magnitude) and spatial (i.e., topography) information on speech processing. Alternatively, the multivariate TMIF determines the multivariate relationship between multiple EEG channels combined and the speech envelope. This latter method is statistically more powerful as it takes interactions between EEG channels into account. However, it

is restricted to temporal interpretations only. For the multivariate TMIF, we used a channel selection including fronto-central and parieto-occipital channels that contribute to speech processing (Lesenfants et al., 2019). Our channel selection is visualized in Supplementary Fig. 1.

Permutation testing

Neural tracking (MI in bits, in this case) is a relative metric and should be compared to a null-distribution to quantify the meaningfulness of the derived values (De Clercq et al., 2023). We created stationary noise that matched the spectrum of the envelope per frequency band individually. Next, we calculated the MI between the noise envelope and the EEG per participant and repeated this process 1000 times. The significance level was then determined as the 95th percentile of permutations per participant (resulting in a single significance level per participant).

No significant differences were found in the significance level between IWA and controls for any frequency band, as determined by Wilcoxon rank sum tests. Supplementary Fig. 2 displays the significance levels of the multivariate TMIF for all frequency bands categorized by group. As no significant differences in the significance level were found between groups, we used a single significance level (i.e., the 95th percentile of permutations across all participants) to interpret the multivariate TMIFs in the Results section.

2.5 Statistics

Group comparisons

We compared neural envelope tracking for IWA with the control group for broadband as well as for delta, theta, alpha, beta and gamma frequency ranges. For the single-channel TMIF, we performed non-parametric spatio-temporal cluster-based permutation tests (Maris and Oostenveld, 2007), indicating clusters in the TMIF over time and space with the largest group difference at threshold $p<0.05$. For the multivariate TMIF, we performed non-parametric temporal cluster-based permutation tests (Maris and Oostenveld, 2007), indicating clusters of samples with the largest group difference at threshold $p<0.05$.

Support Vector Machine Classification

We investigated whether EEG-based envelope tracking outcomes can be used for detecting aphasia. To this end, we used a Support Vector Machine (SVM) to classify held-out participants as control or aphasic using the Scikit-Learn (v. 0.24.2) library in Python (Pedregosa et al., 2011). The multivariate TMIFs for all five individual frequency bands (delta to gamma) were used as input to the model. Additionally, we added age of the participant, as it influences neural envelope tracking (Decruy et al., 2019). We chose a radial basis function kernel SVM and performed a nested cross-validation approach. In the inner

cross-validation, the C-hyperparameter and pruning (i.e., length of the TMIFs) were optimized (accuracy-based) and tested in a validation set using 5-fold cross-validation. The trained model was then tested on the test set, for which we used a leave-one-subject-out cross-validation approach.

The performance of the SVM classifier was evaluated by computing the receiver operating characteristic (ROC) curve and calculating the area under the curve (AUC). We further reported the overall accuracy, the F1-score, the sensitivity and the specificity of the classifier.

Feature contribution. To obtain a proxy for the relevant contribution of each frequency band, we left out a single band and re-fitted the SVM. We repeated this process for all five frequency bands and reported the corresponding performance drop (AUC, accuracy, F1-score).

Recording time

Classification. From a practical perspective, we were interested in how much data the neural envelope tracking technique requires to detect aphasia accurately and obtain stable, reliable results. We iteratively cropped the EEG recording and the envelope in steps of 2 minutes (using the first 1 minute, first 3 minutes, 5, 7... up to the entire 25 minutes of recording time) and calculated the TMIF per frequency band per time duration. Next, we investigated the amount of minutes required for the SVM to reach its classification potential. As described above, we trained and tested our SVM per time duration in the same fashion as the entire duration. Performance (AUC, accuracy, F1-score) was plotted as a function of recording time. We determined the knee point, i.e., the point at which the performance benefit starts to saturate, using the "kneed" python package (Satopaa et al., 2011). The knee point of this curve reflects the point at which the increase in model performance may no longer be worth the corresponding effort.

Within-subjects stability. Second, we investigated the data required to obtain stable, reliable results. We determined the within- and between-subjects stability per time duration. For the within-subjects stability, we individually correlated (Pearson) the TMIF per time duration (i.e., first minute, first 3, 5, ...) with the TMIF of the entire recording per subject. This resulted in a single correlation coefficient for each participant, frequency band and time duration. Next, all correlations were plotted as a function of recording time, and we determined the knee point of the curve on the average across all frequency bands. As such, we gained insight into the amount of data required for a participant's TMIF to become stable (i.e., when there is not much change in an individual's TMIF).

Between-subjects stability. For the between-subjects stability, we calculated each participant's mean MI of the TMIF (integration window 0-400 ms) per time duration (1, 3, 5,... minutes) and the entire recording. This resulted in a single datapoint per participant, frequency band and time duration (i.e., mean MI for a certain duration length x mean MI entire duration). Subsequently, we calculated the correlation coefficient (Pearson's R) between the mean MI for certain time duration and frequency band with the entire recording over participants on the group level, resulting in a single correlation coefficient

per time duration and frequency band. We plotted the correlations as a function of recording time and determined the knee point of the curve on the average across all frequency bands. With this analysis, we investigated the amount of data required for a participant's relative (i.e., compared to other participants) strength of tracking to become stable (i.e., from which point on a participant's relative neural tracking compared to other participants is no longer expected to change).

Split-half reliability

Finally, we report a traditional split-half reliability metric with non-overlapping parts of the recording. We split the EEG recording into two equal parts, i.e., the first 12.5 minutes and the second 12.5 minutes, and computed the TMIFs for each half and each frequency band individually. Next, we computed the mean MI value of the TMIF (0-400 ms) for the first and the second half of the recording per participant individually. Subsequently, we calculated the correlation coefficient (Pearson's R) between the first and second half of the recording over participants on the group level (Pearson's R) for IWA and controls separately.

Data availability statement

We shared our neural tracking outcomes (i.e., the TMIFs) on the Open Science Framework: `https://osf.io/nkmfa/`. Note that our ethical approval does not permit public archiving of raw neuroimaging data, but raw EEG data can be made available upon request and if the GDPR-related conditions are met.

3 Results

3.1 Distinguishing individuals with aphasia from healthy controls

We investigated whether neural envelope tracking is altered in IWA compared to healthy controls. First, we studied the effect in the broadband frequency range (0.5-49 Hz). For the single-channel MI analysis, providing both temporal and spatial information, we found decreased neural envelope tracking for IWA compared to healthy controls (Fig. 1A). A spatio-temporal cluster-based permutation test identified a cluster comprising a large group of fronto-central, parietal and posterior channels (N = 43 channels) from 0.11 s to 0.3 s (p=0.004), centered around the second peak. The multivariate MI analysis, which combines information from multiple channels, confirmed these results: a temporal cluster-based permutation test identified a cluster between 0.11 s to 0.26 s in which IWA displayed a decreased response (p = 0.005) (Fig. 1B).

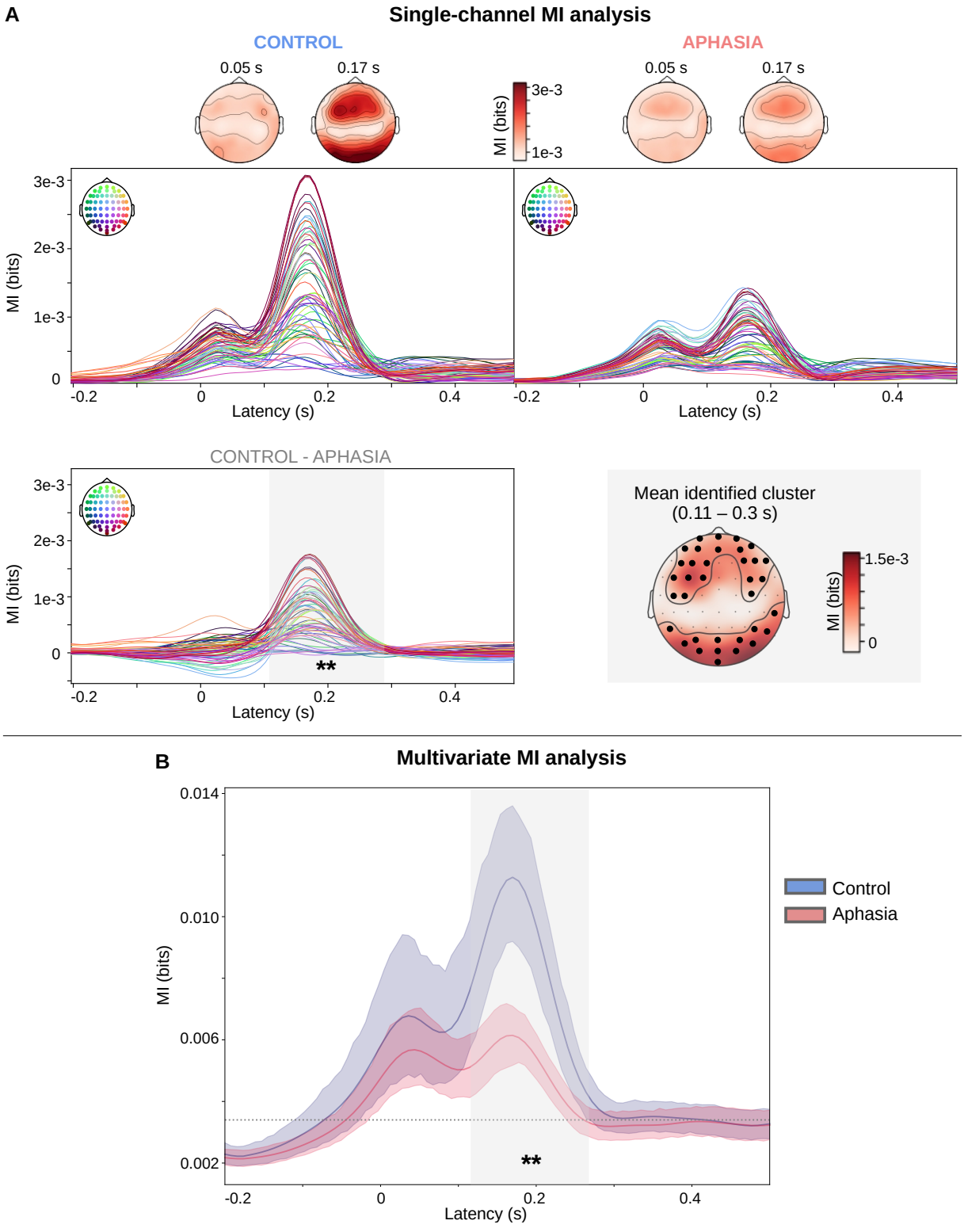


Figure 1. Broadband analysis. A. The average single-channel TMIF for the control and the aphasia group separately, with topoplots at the first and second peak (0.05 and 0.17 s). The spatio-temporal cluster-based permutation test investigated the difference between the control and aphasia group (control - aphasia) and identified a cluster (below threshold $p<0.05$) with the largest group difference, centered around the second peak. Brain latencies belonging to the cluster are marked in a shaded gray area, the channels belonging to the cluster are indicated with a black dot on the topoplot. **B.** The group average TMIF, for both groups separately. The shaded, colored areas indicate the 95% confidence interval. The shaded gray area indicates the cluster with largest group difference (threshold $p<0.05$), identified using a temporal cluster-based permutation test. ** = $p<0.01$

We further investigated the neural response in narrow frequency bands. We focused on the multivariate TMIF as it is a statistically more robust method compared to the single-channel TMIF, and we used those features as input to our SVM classifier in the subsequent section. The single-channel TMIFs for all frequency bands are provided in the Supplementary materials. We generally observed decreased neural envelope tracking for IWA compared to healthy controls (Fig. 2). Temporal cluster-based permutation tests identified clusters below threshold $p<0.05$ for delta (0.1 to 0.30s, $p=0.003$), theta (0.04 to 0.27s, $p=0.005$) and gamma (0.01 to 0.1s, $p = 0.004$) band. No clusters exceeding the $p<0.05$ threshold were detected for the alpha and beta bands. These results are confirmed in the single-channel MI analysis, where spatio-temporal cluster-based permutation tests identified clusters for delta, theta and gamma band for a large group of fronto-central, parietal and posterior channels (visualizations and statistics provided in the Supplementary materials).

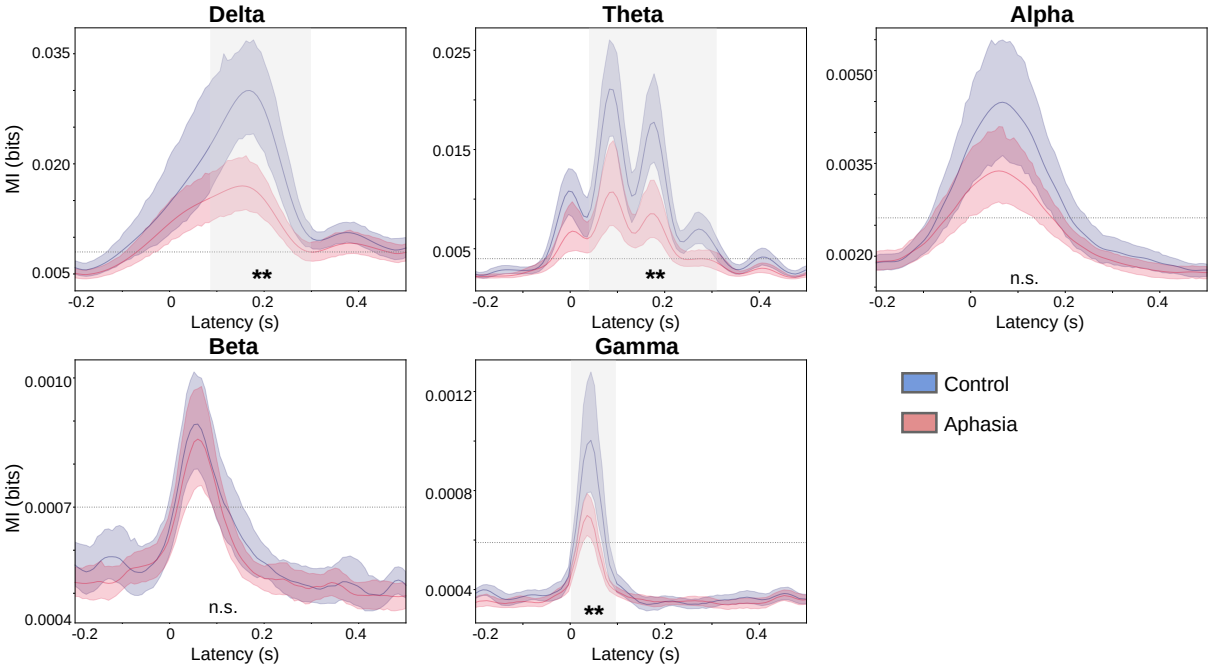


Figure 2. Frequency-specific analysis. Group average TMIF's visualized per frequency band, with colored shaded areas indicating the 95% confidence interval. Shaded, gray areas indicate clusters with largest group difference (below threshold $p<0.05$) identified using temporal cluster-based permutation tests. ** = $p<0.01$

3.2 Support Vector Machine classification

Next, we investigated whether we could detect aphasia based on neural envelope tracking measures in the individual frequency bands. To this end, we used an SVM to classify participants as belonging to the aphasia or the healthy control group via leave-one-out cross-validation. We used the TMIFs in our five frequency bands of interest and age as input features to the model. The SVM successfully classified participants belonging to either group with an accuracy of 83.67%, an F1-score of 83.58% and an AUC of 88.05%. The SVM had a sensitivity of 88.89% and a specificity of 77.27% for aphasia. Fig. 3A displays

the ROC curve.

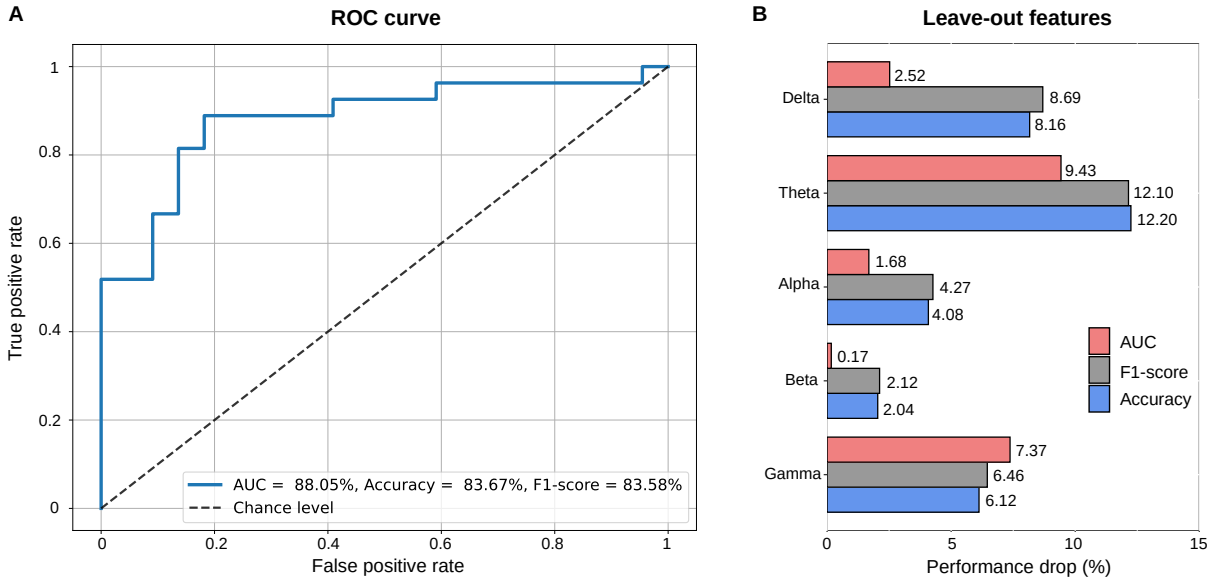


Figure 3. Result of the SVM classifier. A. Receiver operating characteristic curve (ROC). **B.** Relative feature importances. Drop in performance is visualized after leaving out the corresponding feature (i.e., frequency band).

To obtain a measure of relative frequency band contribution, we iteratively left out a frequency band and trained the SVM with the remaining features. For each left-out frequency band, we calculated the performance drop. As assessed with accuracy and F1-score metrics, theta, followed by delta, gamma, alpha and beta caused the largest drop in performance (see Fig. 3B). When estimated with AUC, the accuracy was still the highest for theta, followed by gamma, delta, alpha and beta. This confirmed our group comparison analyses: delta, theta and gamma band are the most relevant, discriminating features.

3.3 Recording length

We further investigated how much data the neural envelope tracking technique requires for robust and stable results (Fig. 4). With only one minute of recording time, the SVM obtained classification accuracy close to chance-level (55%). Yet, from 5 minutes on, the SVM reached an accuracy of 81.63%, and performance fluctuated between 81.63% and 85.71% for the remaining part of the recording. In practice, this corresponds to one less or one additional correctly classified participant with respect to the full recording (Fig. 3A). The knee point of the curve was identified at 5 minutes of recording length. From 9 minutes on, the SVM converged to an AUC of 80%. However, compared to the entire recording length (AUC=88.05%), its full potential is reached from 13 minutes on (AUC robustly crossed 85%, with a maximum of 89.73% at 15 minutes). The F1-score mostly overlapped with the accuracy and never differed more than 0.44% . The SVM performance is plotted as a function of recording length, displayed in Fig. 4A.

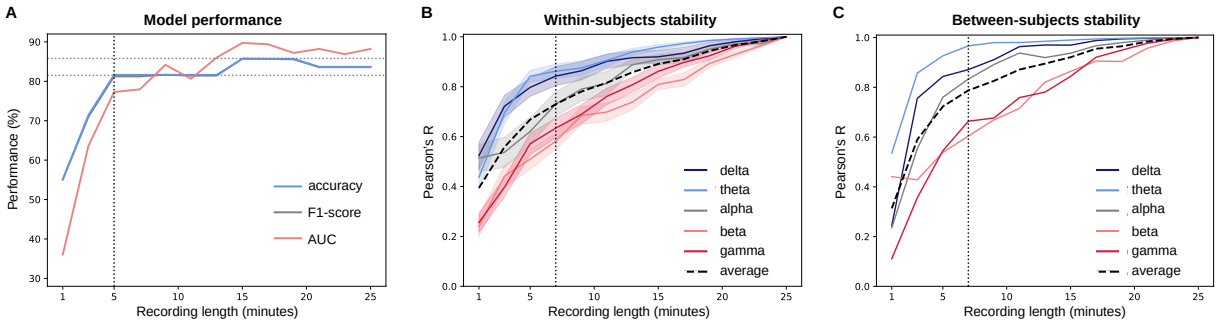


Figure 4. Recording Length. **A.** Performance (accuracy, F1-score, AUC) of the SVM classifier plotted as a function of time. **B.** Within-subjects stability for all five frequency bands and the average across frequency bands (black line). Shaded areas indicate the standard error. **C.** Between-subjects stability for all five frequency bands and the average across frequency bands (black line). The knee point of all panels is indicated with a vertical dotted line (based on the average for panels **B** and **C**).

The within-and between-subjects stability is plotted as a function of recording time in Fig. 4B and 4C. Highest within- and between-subjects correlations were observed for the low-frequency bands, namely delta and theta. Taking the average of all frequency bands, we identified the curve's knee point at 7 minutes of recording length (see black dotted lines). The within-subjects stability (Fig. 1 4B) had an average correlation of R=0.73, and the between-subjects stability (Fig. 4C) had an average correlation of R=0.79 at the knee point of the curve.

3.4 Reliability of neural envelope tracking

Finally, we calculated the split-half reliability of neural envelope tracking. Table 1 provides the correlations and statistics. We generally found higher correlations in the lower frequency bands (delta and theta). Correlations were comparable between IWA and the control group; the 95% confidence intervals overlapped for each frequency band, and post-hoc Fisher z-tests, performed using the 'cocor' package in Rstudio (Diedenhofen and Musch, 2015), revealed no difference in correlation strength between both groups.

Table 1. Split-half reliability.

	delta		**theta**		**alpha**		**beta**		**gamma**	
	IWA	C	IWA	C	IWA	C	IWA	C	IWA	C
Pearson's R	0.85	0.91	0.96	0.92	0.75	0.88	0.61	0.62	0.62	0.73
CI	[0.68 ; 0.93]	[0.78 ; 0.96]	[0.91 ; 0.98]	[0.82 ; 0.96]	[0.52 ; 0.88]	[0.72 ; 0.95]	[0.31 ; 0.81]	[0.27 ; 0.83]	[0.31 ; 0.81]	[0.43 ; 0.88]
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.006	0.02	0.007	0.002
Fisher's z (p-value)	z=0.85 (1)		z=-0.94 (1)		z=1.24 (1)		z=0.04 (1)		z=0.66 (1)	

CI=95% confidence interval; C=Control group. Fisher z-test comparing controls - IWA. All p-values are corrected for multiple comparisons.

4 Discussion

We conducted an in-depth study on neural envelope tracking of natural speech in post-stroke aphasia. First, we found that IWA display decreased neural envelope tracking compared to heatlhy controls for a broadband frequency range. Second, frequency-specific analyses indicated that group differences are most prominent in the delta, theta and gamma frequency ranges. Third, the suitability of neural envelope tracking measures as a biomarker for post-stroke aphasia was demonstrated using an SVM classifier which yielded high accuracy (84%, AUC 88%). Finally, we showed that an assessment based on neural envelope tracking could be obtained in a time-efficient (5 minutes of EEG recording) and highly reliable manner.

4.1 Individuals with aphasia display decreased neural envelope tracking

Broadband frequency analysis

Neural envelope tracking at broadband is decreased in IWA compared to healthy controls. The single-channel TMIF analysis revealed a cluster at neural response latencies centered around the second peak in the TMIF comprising a large group of fronto-central, temporal and parieto-occipital channels (see Fig. 1A). The multivariate TMIF confirmed this result: a temporal cluster comprising brain latencies surrounding the second peak in the TMIF (Fig. 1B) was identified. A recent neural tracking study showed that the second peak emerges when speech is comprehensible and diminishes when it is not understood. By contrast, the first peak displayed a prominent response when speech was incomprehensible (Verschueren et al., 2022). Thus, the second peak we observed here is most likely related to speech understanding, while the first peak is likely more implicated in acoustically processing the signal. Therefore, it is not surprising that in IWA, where language understanding is impaired, the second peak in the TMIF is decreased compared to healthy controls.

Frequency-specific analysis

We further investigated neural envelope tracking in narrow frequency bands. Our findings revealed a decrease in tracking for IWA compared to healthy controls in the low-frequency bands (delta and theta, see Fig. 2A and 2B), which are crucial for speech understanding (Vanthornhout et al., 2018). The delta band encodes sentences, phrases and words (Kaufeld et al., 2020; Keitel et al., 2018), while theta band tracks the syllable rate of the stimulus (Etard and Reichenbach, 2019; Lizarazu et al., 2019). Neural tracking in the low-frequency bands drops when these linguistic units become incomprehensible (Kaufeld et al., 2020; Keitel et al., 2018; Xu et al., 2022). Atypical neural tracking of the low-frequency temporal envelope has been reported in several clinical populations, including individuals with primary progressive aphasia (Dial et al., 2021) and dyslexia (Di Liberto and Lalor, 2017; Lizarazu et al., 2021; Mandke et al., 2022). In the case of dyslexia, which is characterized by phonological processing difficulties, alterations in

low-frequency envelope tracking are believed to reflect an atypical sampling mechanism that affects faster modulations at the phoneme and grapheme level (Mandke et al., 2022). These findings in healthy and clinical populations demonstrate the potential of neural tracking measures of the low-frequency envelope as a biomarker for language impairments.

Fewer studies investigated the role of high-frequency neural envelope tracking. Some studies suggest a role for alpha and beta in attention and auditory-motor coupling (Wöstmann et al., 2017; Fujioka et al., 2015), and for the gamma band in encoding phonetic features (Hyafil et al., 2015; Giraud and Poeppel, 2012; Gross et al., 2013). Our study found no group differences in the alpha and beta bands. However, individuals with aphasia displayed decreased neural envelope tracking in the gamma band. The neural response in the gamma band was characterized by an early response peak (Fig. 2) and a group difference present in the right hemisphere (see Supplementary Fig. 7). This early response latency in the gamma band aligns with the idea of a linear phase property, where the neural response delay in higher frequency bands is shorter (Zou et al., 2021). A somewhat similar neural response pattern characterized by an early response latency and a right hemisphere bias in the high gamma band (>70 Hz) was also reported by Kulasingham et al. (2020). The gamma band has been of particular interest in dyslexia research, with several studies reporting alterations in gamma band activity. However, most studies have focused on phase-locking and phase coherence in response to amplitude-modulated noise (for a review, see (Lizarazu et al., 2021)), while few have investigated gamma band neural envelope tracking during natural speech processing (Mandke et al., 2022). To gain a better understanding of (low-)gamma band neural envelope tracking of natural speech, which we have shown to demonstrate robust group differences and contribute significantly to detecting aphasia (as depicted in Fig. 3B), future research should aim to further investigate its implications for speech understanding and language impairments.

In line with the idea that individual frequency bands are involved in different speech processes, exploratory analysis revealed low to moderate positive and negative correlations between neural tracking in individual frequency bands (see Supplementary Table 1). This suggests that a participant with high neural tracking in one frequency band may not necessarily display high neural tracking in other frequency bands. In contrast, a broadband frequency analysis shows high redundancy compared to the delta band (R=0.79 for IWA, R=0.92 for controls), which can be attributed to the fact that most power in the EEG and the envelope is concentrated in the lowest frequencies. This highlights the relevance of conducting frequency-specific analyses. We believe that the use of frequency-specific features to train the SVM favored good classification results, as discussed in the next section. Future research should investigate whether the neural response to these frequency bands may capture specific language deficits.

4.2 High accuracy detection of post-stroke aphasia

We assessed the suitability of the neural tracking technique to detect post-stroke aphasia using an SVM classifier with the TMIFs computed in the individual frequency bands as input to the model. The SVM robustly detected aphasia, with an accuracy of 83.67%, an F1-score of 83.58% and an AUC of 88.05%. The ROC curve, plotting the true positive as a function of false positive aphasia classification, is depicted in Fig. 3A. The relative contribution of individual frequency bands for detecting aphasia at the individual level confirmed our group comparison results: delta, theta and gamma band neural tracking were most predictive for capturing aphasia (see Fig. 3B).

These performance outcomes of the SVM can be interpreted against behavioral assessment. As described in the Methods section, 7 out of 27 IWA (i.e., 26%) did not score below the cut-off threshold on either of the two diagnostic language tests for aphasia administered during the study El Hachioui et al. (2017); Van Ewijk et al. (2020). Nevertheless, these subjects had extended language deficit documentation and followed speech-language therapy at the time of participation. Although a more extensive screening for aphasia could have identified a language deficit, this finding highlights the challenge of detecting aphasia in the chronic phase following a stroke. While further investigation is required, the higher detection accuracy of the EEG-based neural tracking classification suggests that it may be more sensitive than behavioral screening tests for capturing subtle language problems in individuals with aphasia.

Nonetheless, comparing the SVM classification accuracy to behavioral assessment is rather difficult, as the underlying tested language skills are different. Standardized aphasia tests use isolated sounds, phonemes, words or short sentences, questioning the ecological validity of such tasks (Hamilton and Huth, 2018). Consequently, research reports a discrepancy between common test outcomes and everyday life speech assessments (Lesser and Algar, 1995; Kim et al., 2022; Stark et al., 2021; Wallace and Kimelman, 2013), and cognitive problems can bias the test result (Fonseca et al., 2019; Rohde et al., 2018). While there exist behavioral natural speech assessments (Armstrong, 2000), they are only limitedly applied in practice due to the high workload (time-intensive) and a lack of knowledge of natural speech analyses (Bryant et al., 2019; Stark et al., 2021). Novel automatic speech recognition and natural language processing techniques (Dalton et al., 2022; Jamal et al., 2017; Le et al., 2018) may provide a solution in the future. The neural tracking technique directly addresses the limitation of low ecological validity from which behavioral tests suffer.

4.3 Assessing time-efficiency, stability and reliability

We investigated how much data neural envelope tracking requires to detect aphasia accurately and yield reliable results. We assessed SVM classification performance as a function of recording time, as shown in Fig. 4A. Our findings indicate that high-accuracy detection can be achieved with just 5 minutes of recording time (accuracy of 81.63%). However, extending the recording duration to 13 minutes can

provide additional benefits in terms of the AUC, with a robust increase above 85%, and a maximum AUC of 89.73% achieved at 15 minutes. In summary, our results demonstrate that neural envelope tracking can effectively detect aphasia in a time-efficient manner, consistent with prior recommendations that language assessments in aphasia should not exceed 15 minutes to avoid fatigue and cognitive/attentional challenges (El Hachioui et al., 2014). Our findings have important implications for potential clinical applications of neural envelope tracking in aphasia.

We further investigated the amount of data necessary for our frequency band features to achieve stability. Our within- and between-subjects stability analysis revealed that 7 minutes recording length is sufficient for the TMIF of individual subjects and at the group-level to resemble the TMIF from the entire recording (see Fig. 4B and 4C). These findings were consistent within both groups, with 7 minutes being the minimum recording length required (as illustrated in Supplementary Fig. 8). Notably, we observed that lower frequency bands (delta and theta) converge more rapidly compared to higher frequency bands. Within 3-5 minutes, stability correlations for these bands were relatively high, robustly crossing R=0.80. In contrast, higher frequency bands require a longer time to converge and exhibit a more linear slope compared to the delta and theta bands. These less stable results and longer minimal recording length for the higher frequency bands can be attributed to their lower signal-to-noise ratio.

To summarize, our results indicate that 5-7 minutes of recording time are sufficient for assessing neural envelope tracking at low-frequency ranges, which reflect higher-level linguistic processes and speech understanding. However, a more comprehensive evaluation that includes higher frequency bands, which can provide minor additional benefits, requires a slightly longer recording duration (>13 minutes). These findings are consistent with previous research in healthy participants, which suggests that low-frequency neural tracking requires approximately 3-10 minutes of recording time for robust outcomes (Desai et al., 2023; Di Liberto and Lalor, 2017; Mesik and Wojtczak, 2022). Our study contributes an innovative approach by defining the minimal recording length required to detect language impairments at the individual level and suggests that the recording duration for future studies in individuals with language impairments may depend on the specific research question being addressed.

Previous studies on aphasia using ERPs have suggested the potential of this approach for clinical diagnosis, but without reporting on its reliability (Cocquyt et al., 2020). However, evaluating the reliability of test results is crucial to determine the usefulness of capturing individual language impairments. In this study, we assessed the reliability of neural envelope tracking using split-half reliability metrics. The results demonstrate strong correlations between both halves, particularly in the delta and theta bands (Table 1). Our findings are consistent with previous research reporting a correlation of R=0.89 for delta and R=0.82 for theta across two stories in a cohort with language impairments caused by a neurodegenerative disorder (Dial et al., 2021). Yet, reliability measures in our study were generally lower for higher frequency bands. As mentioned earlier, these bands require more data to converge and have a

lower signal-to-noise ratio. It is worth noting that our reliability measure may be affected by fatigue. Thus, future studies should examine the generalizability of the results across stories and speakers at different sessions (i.e., test-retest) to further investigate the reliability of neural tracking for applications in aphasia.

4.4 Limitations and future directions

Our study demonstrates that neural envelope tracking is a reliable and accurate method for detecting language impairments in aphasia. However, our current approach does not provide information on the specific language profile of the patient (i.e., which underlying language component, e.g., auditory, phonetic, semantic,... is affected). Investigating these deficits would require a larger sample size with a more uniform spread of aphasia severity levels. In future research, we suggest exploring whether neural tracking in specific frequency bands can cluster different language profiles in aphasia. In addition, recent studies investigated the neural response to speech representations beyond the temporal envelope. For example, it has been shown that linguistic speech representations at phoneme and word level can improve the model's fit to the EEG (Di Liberto et al., 2015; Gillis et al., 2021) and can provide complementary information on speech processing (Verschueren et al., 2022; Gillis, Kries et al., 2023). Future research should (1) examine whether incorporating these linguistic speech representations can enhance aphasia detection and inform on specific language deficits and (2) assess the reliability and robustness of these features, which is currently lacking in the literature.

Several other open questions must be addressed before neural tracking can be applied in clinical settings. Firstly, neural tracking must be applied to IWA in the acute stage after stroke. This work considered the chronic stage only since it is characterized by a more stable language profile (Johnson et al., 2019). Secondly, the present study distinguished IWA and healthy controls only. If neural tracking is to be used for screening aphasia in the acute stage post-stroke, a clear dissociation between stroke patients with and without aphasia is crucial. However, such dissociation is generally not considered in behavioral screening tests despite being used in the clinic on a daily basis (Rohde et al., 2018). Lastly, this study used language stimuli in the receptive domain only. Recent studies have suggested that the same analysis can also be applied to the expressive domain, i.e., speech production (Perez et al., 2022), which could open new perspectives to studying expressive language problems in IWA.

Conclusion

This study investigated neural envelope tracking of natural speech in patients with chronic post-stroke aphasia. The findings showed that individuals with aphasia exhibited reduced brain responses in the delta, theta, and gamma bands, likely reflecting decreased processing of higher-level auditory and linguistic units. The study also demonstrated the efficacy of neural tracking in capturing language impairments

at the individual level in a highly reliable and time-efficient manner, which suggests its promising clinical potential as an assessment tool. Despite these positive results, several open questions remain that need to be addressed before neural tracking can be used in clinical settings. For instance, it remains unclear whether neural tracking can accurately capture specific language problems, and its effectiveness in assessing patients in the acute stage post-stroke requires future investigation. Nevertheless, our work represents a significant step towards more automatic and ecologically valid assessments of language problems in aphasia.

Acknowledgements

The authors would like to express their heartfelt gratitude to all the participants, particularly those with aphasia and their families that supported them. We would also like to extend our thanks to Dr. Klara Schevenels for her assistance in the recruitment process, as well as the individuals that helped with the data collection: Janne Segers, Rosanne Partoens, Charlotte Rommel, Ines Robberechts, Laura Van Den Bergh, Anke Heremans, Frauke De Vis, Mouna Vanlommel, Naomi Pollet, Kaat Schroeven, Pia Reynaert and Merel Dillen.

Funding

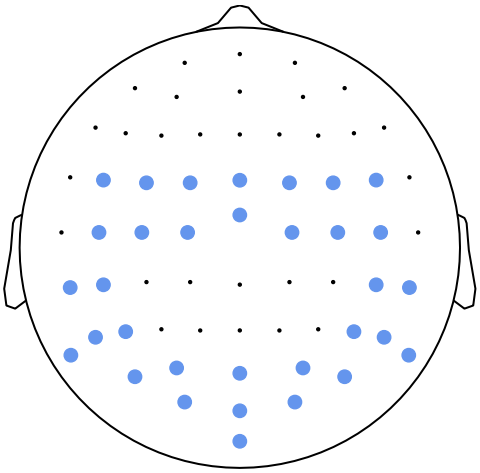
Research of Pieter De Clercq was supported by the Research Foundation Flanders (FWO; PhD grant 1S40122N). Jill kries was financially supported by the Luxembourg National Research Fund (FNR; AFR-PhD project reference 13513810). Research of Jonas Vanthornhout was supported by FWO (postdoctoral grant: 1290821N). The presented study further received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Tom Francart; grant agreement No. 637424), and by the FWO grant No. G0D8520N.

Competing interests

The authors declare no conflicts of interest, financial or otherwise.

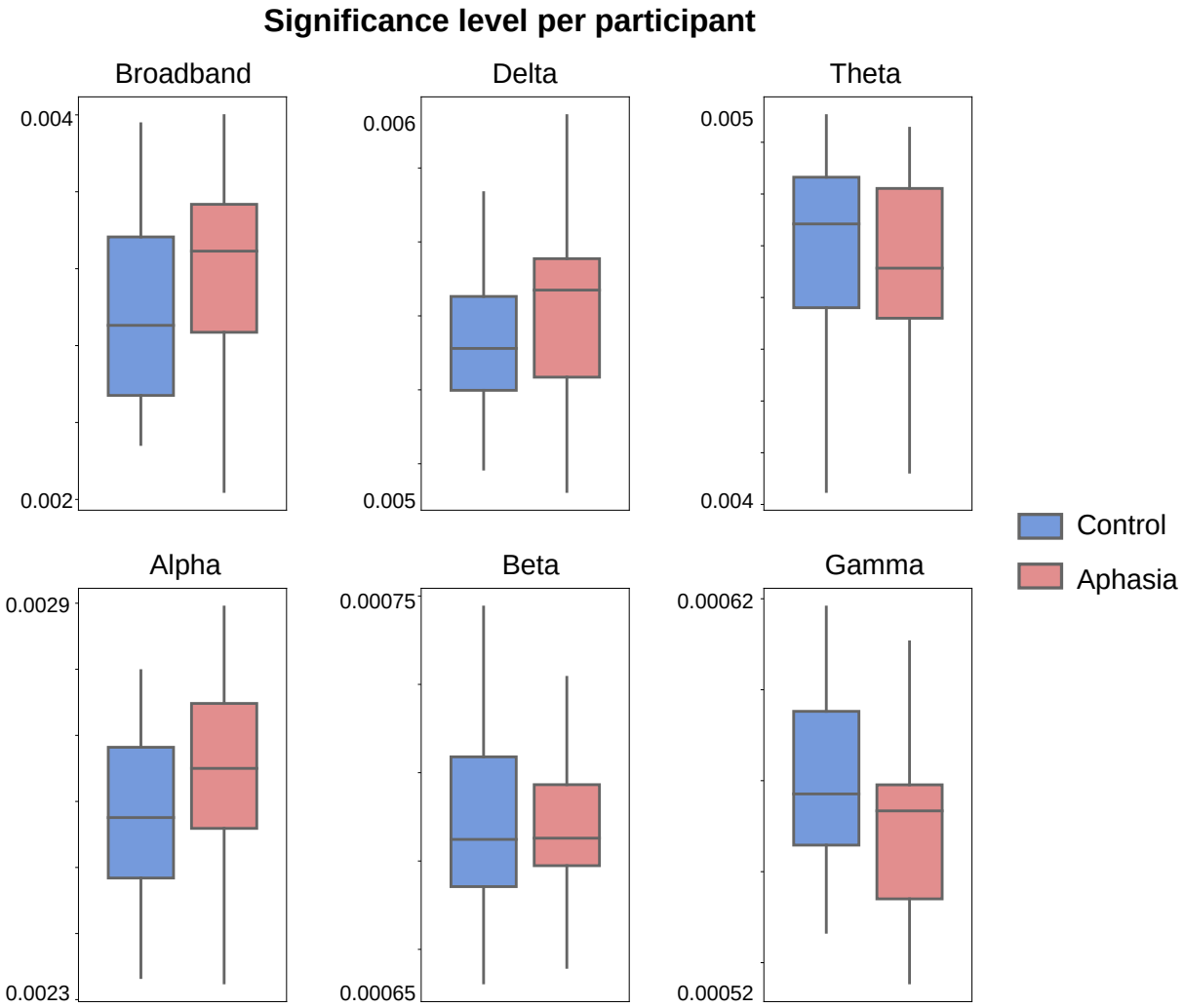
Supplementary material

Channel Selection



Supplementary Fig. 1. Channel selection.

Significance level of neural envelope tracking

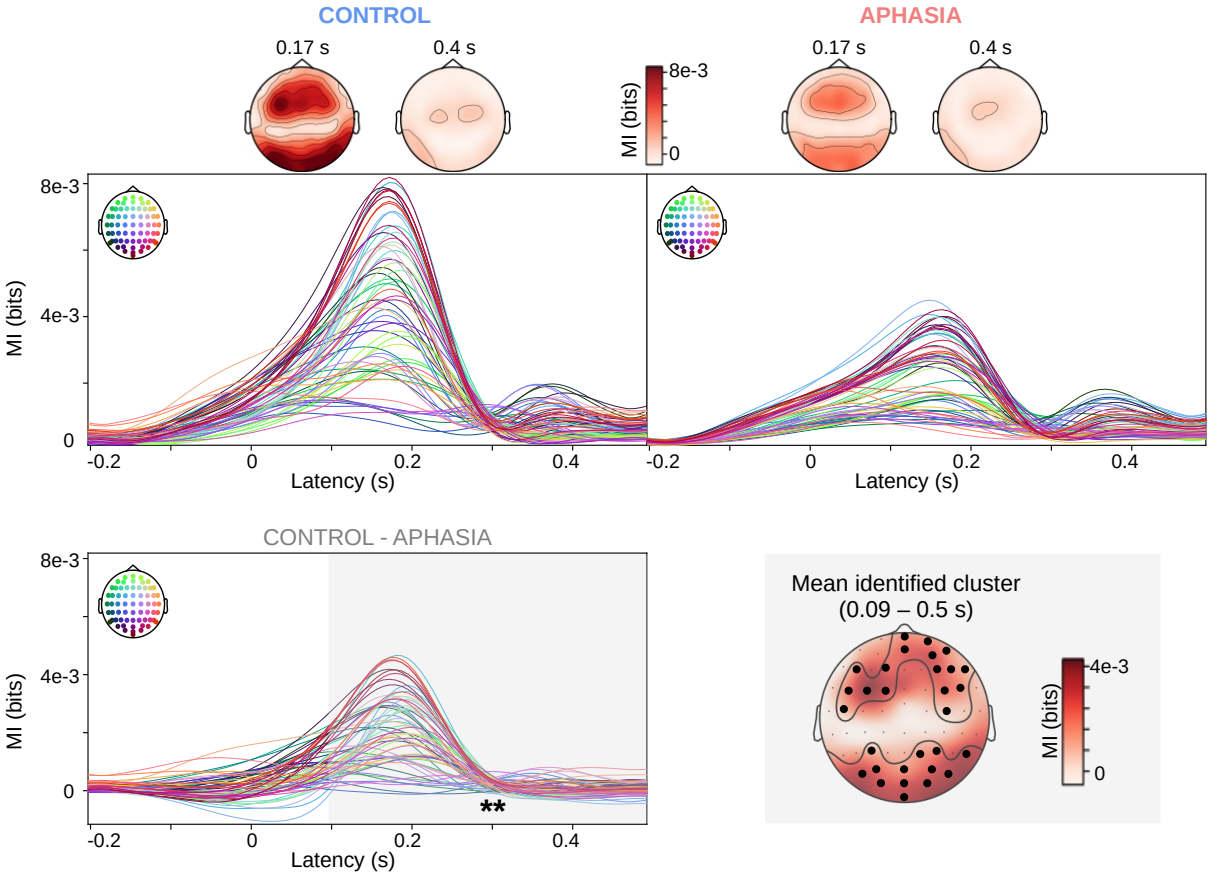


Supplementary Fig. 2. Significance level of neural tracking. Boxes represent the 95th percentile of permutations per subject and per frequency band. There was no significant difference between groups for any frequency band.

Single-channel TMIF analysis

Delta band

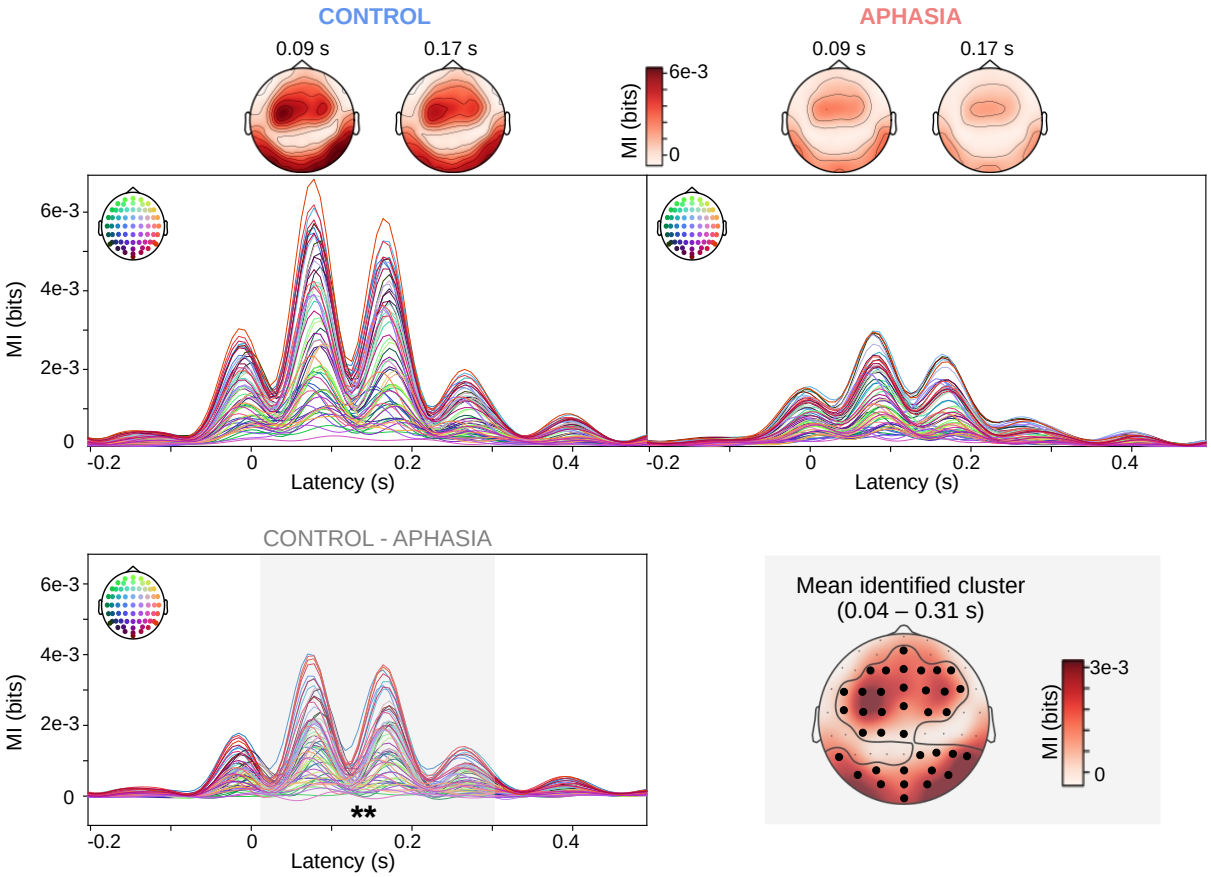
The single-channel TMIF analysis revealed decreased delta band envelope tracking for IWA compared to healthy controls. A spatio-temporal cluster-based permutation test identified a cluster (p=0.005) comprising a large group of bilateral fronto-central, parietal and posterior channels (N = 46 channels) and brain latencies from 0.09 s to 0.5 s. The results are depicted in Fig. 3.



Supplementary Fig. 3. Delta band analysis. The average single-channel TMIF in delta band for the control and the aphasia group separately, with topoplots at indicated brain latencies. The spatio-temporal cluster-based permutation test investigated the difference between the control and aphasia group (control - aphasia) and identified a cluster (below threshold p<0.05) with the largest group difference. Brain latencies belonging to the cluster are marked in a shaded gray area, the channels belonging to the cluster are indicated with a black dot on the topoplot.** = p<0.01

Theta band

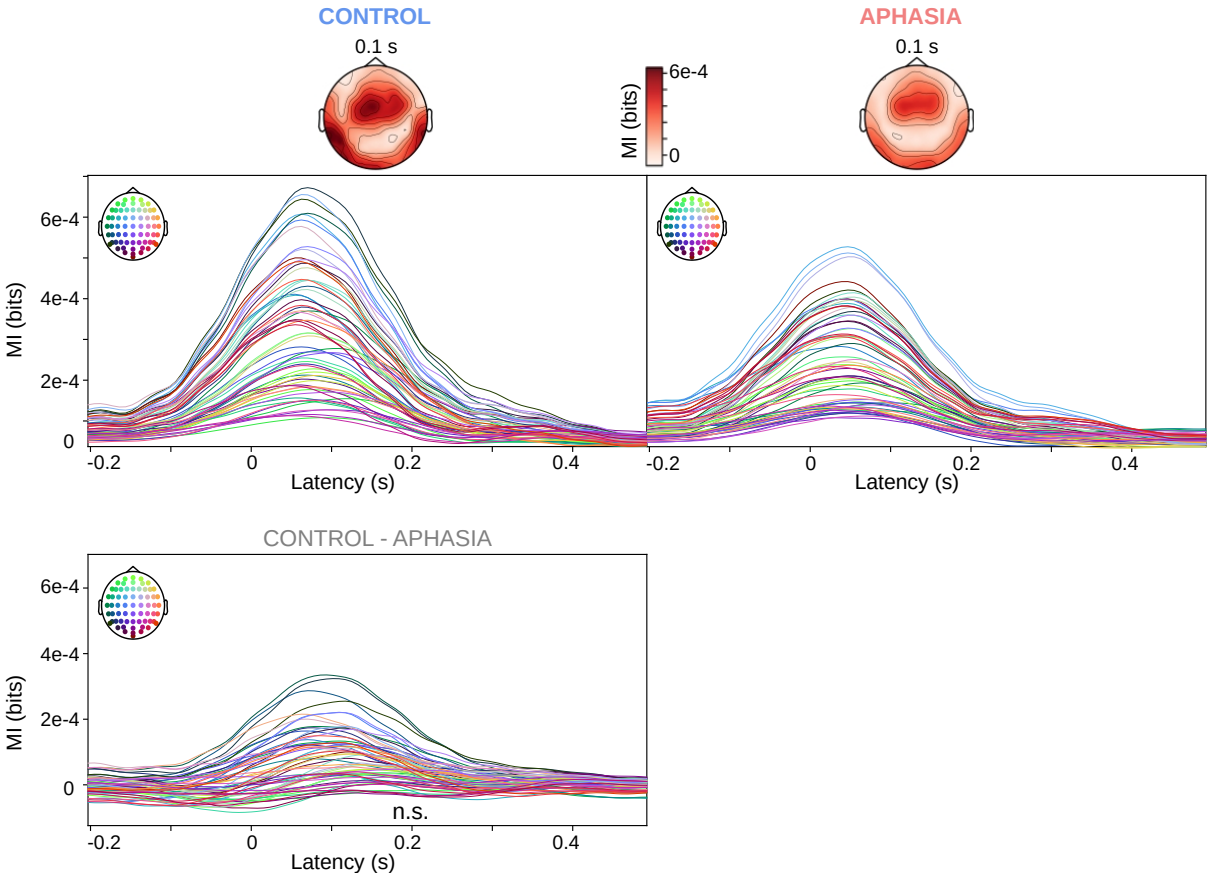
For the theta band, the single-channel TMIF analysis revealed decreased envelope tracking for IWA compared to healthy controls. A spatio-temporal cluster-based permutation test identified a cluster (p=0.005) comprising a large group of bilateral fronto-central, parietal and posterior channels (N = 40 channels) and brain latencies from 0.09 s to 0.31 s. Fig. 4 visualizes the result.



Supplementary Fig. 4. Theta band analysis. The average single-channel TMIF in theta band for the control and the aphasia group separately, with topoplots at indicated brain latencies. The spatio-temporal cluster-based permutation test investigated the difference between the control and aphasia group (control - aphasia) and identified a cluster (below threshold $p<0.05$) with the largest group difference. Brain latencies belonging to the cluster are marked in a shaded gray area, the channels belonging to the cluster are indicated with a black dot on the topoplot.** = $p<0.01$

Alpha band

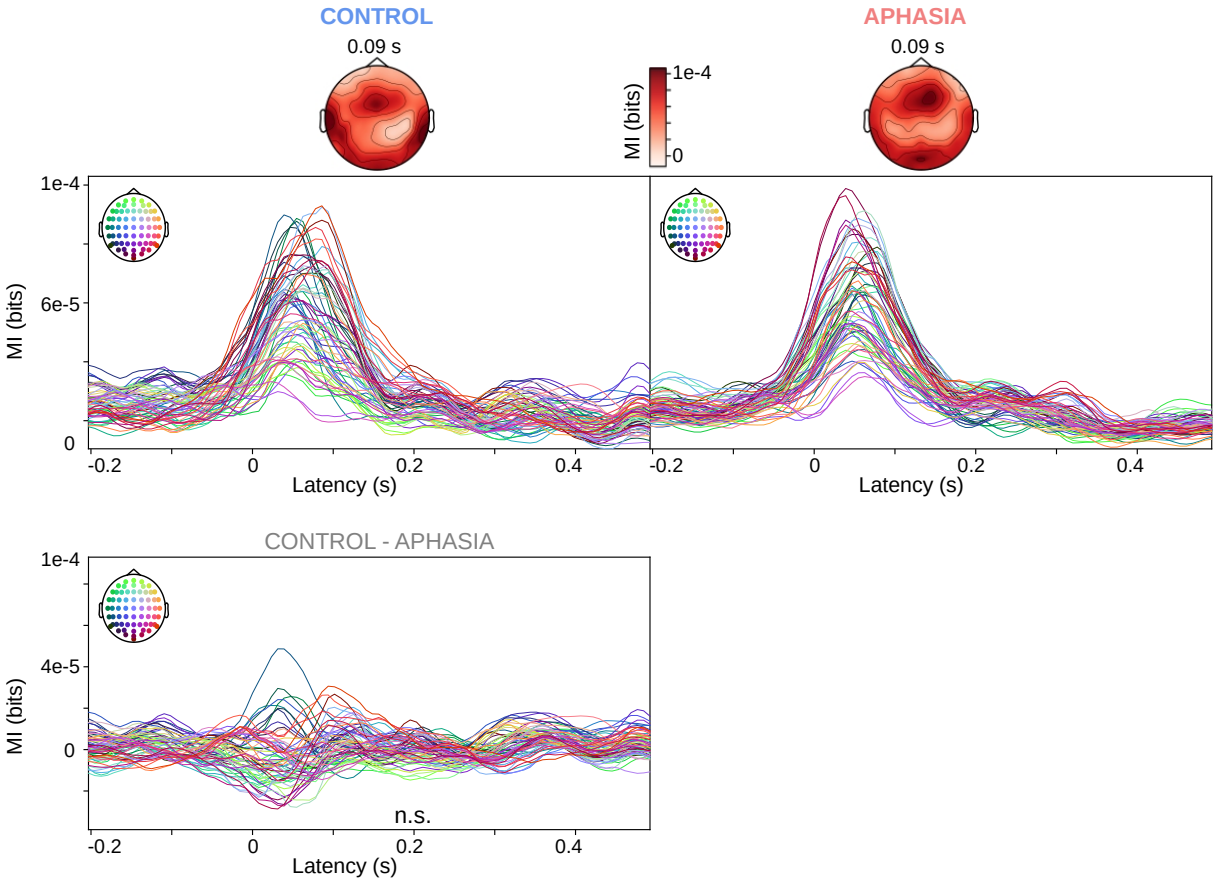
In the alpha band, a spatio-temporal cluster-based permutation test found no clusters exceeding $p<0.05$ threshold level. The group results are displayed in Fig. 5.



Supplementary Fig. 5. Alpha band analysis. The average single-channel TMIF in alpha band for the control and the aphasia group separately, with topoplots at indicated brain latencies. The spatio-temporal cluster-based permutation test investigated the difference between the control and aphasia group (control - aphasia), but did not find a group difference with p-value below threshold level 0.05.

Beta band

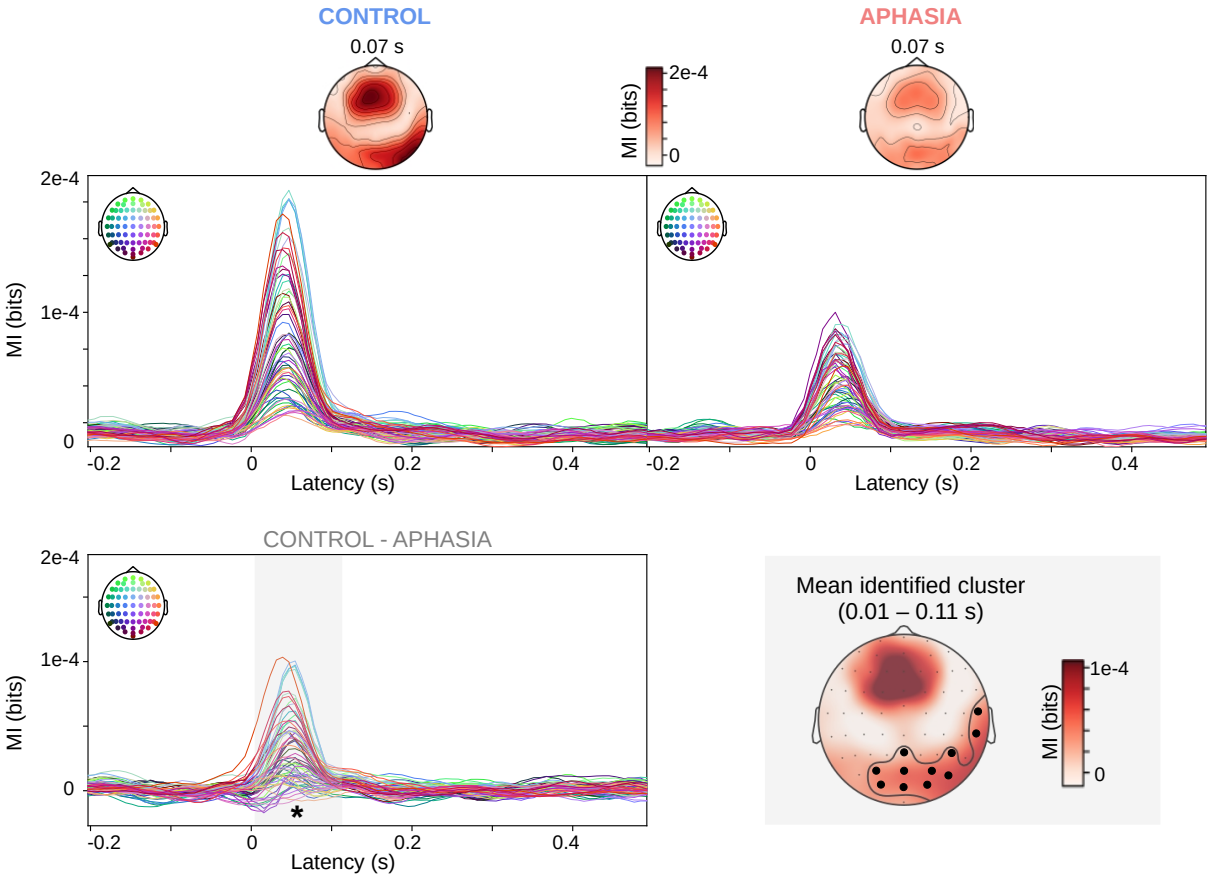
In the beta band, a spatio-temporal cluster-based permutation test found no clusters exceeding $p<0.05$ threshold level. The group results are displayed in Fig. 6.



Supplementary Fig. 6. Beta band analysis. The average single-channel TMIF in beta band for the control and the aphasia group separately, with topoplots at indicated brain latencies. The spatio-temporal cluster-based permutation test investigated the difference between the control and aphasia group (control - aphasia), but did not find a group difference with p-value below threshold level 0.05.

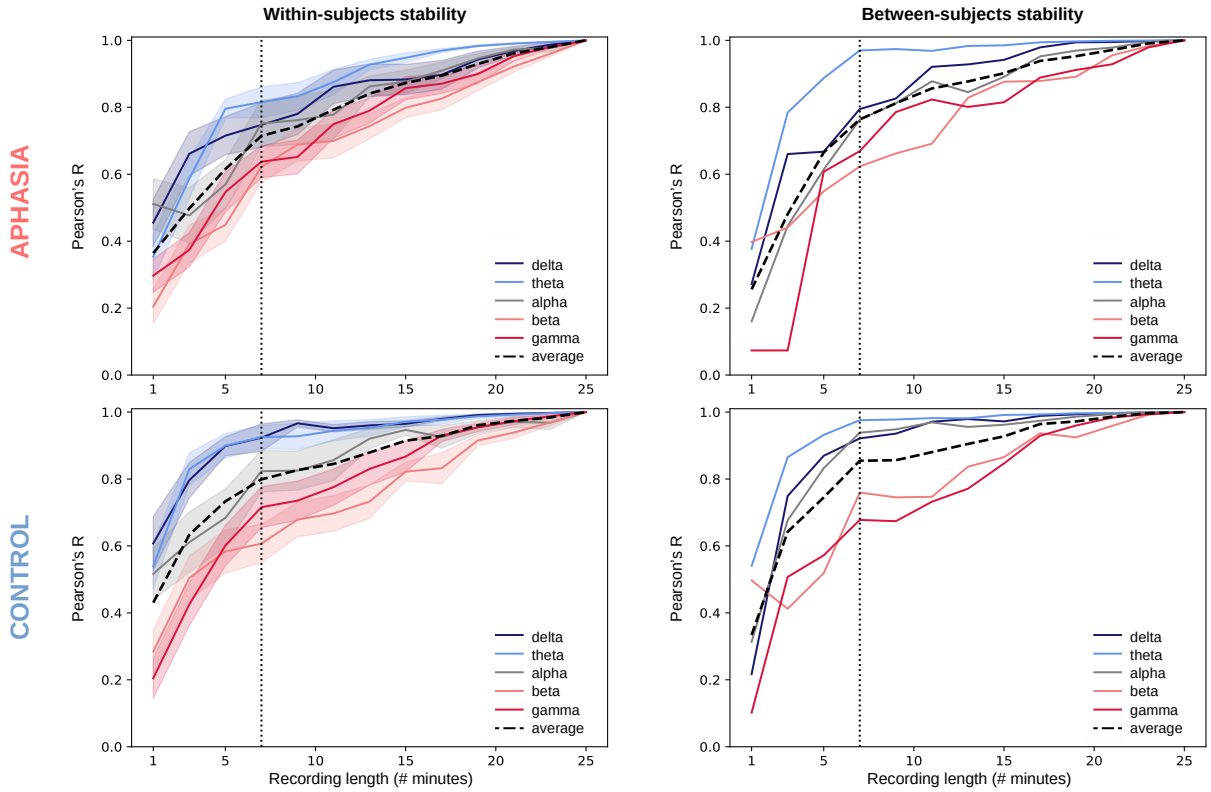
Gamma band

Finally, IWA displayed decreased neural envelope tracking in the gamma band. A spatio-temporal cluster-based permutation test identified a cluster (p=0.03) comprising parietal and posterior channels (N = 12 channels), primarily in the right hemisphere, and brain latencies from 0.01 s to 0.11 s. Fig. 7 visualizes the result.



Supplementary Fig. 7. Gamma band analysis. The average single-channel TMIF in gamma band for the control and the aphasia group separately, with topoplots at indicated brain latencies. The spatio-temporal cluster-based permutation test investigated the difference between the control and aphasia group (control - aphasia) and identified a cluster (below threshold p<0.05) with the largest group difference. Brain latencies belonging to the cluster are marked in a shaded gray area, the channels belonging to the cluster are indicated with a black dot on the topoplot.* = p<0.05

566 # Group-specific stability analysis



Supplementary Fig. 8. Stability measures grouped. Within- and between-subjects stability analysis performed for each group separately. Black dotted line indicates the average across frequencies. Shaded areas indicate the standard error of the correlations. The knee point of all panels is indicated with a vertical dotted line (based on the average across frequencies)

Correlation matrix frequency bands

Supplementary Table 1. Correlation matrix neural envelope tracking

	IWA						Controls					
	broad	delta	theta	alpha	beta	gamma	broad	delta	theta	alpha	beta	gamma
broad	1						1					
delta	0.79	1					0.92	1				
theta	0.15	0.01	1				0.09	0.02	1			
alpha	0.15	0.07	0.71	1			0.27	0.15	0.52	1		
beta	0.15	0.20	0.62	0.61	1		0.48	0.57	0.32	0.44	1	
gamma	0.19	0.22	-0.01	0.01	0.23	1	0.12	0.18	-0.18	-0.22	0.26	1

Exploratory analysis investigating the collinearity between frequency bands. The table displays the Pearson correlations for the mean MI (integration window 0-400 ms).

References

Aerts, A., van Mierlo, P., Hartsuiker, R. J., Santens, P., and Letter, M. D. (2015). Neurophysiological sensitivity for impaired phonological processing in the acute stage of aphasia. *Brain and Language*, 149:84–96.

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, 14(9):875–892.

Becker, F. and Reinvang, I. (2007). Successful syllable detection in aphasia despite processing impairments as revealed by event-related potentials. *Behavioral and Brain Functions*, 3.

Brodbeck, C., Das, P., Gillis, M., Kulasingham, J. P., Bhattasali, S., Gaston, P., Resnik, P., and Simon, J. Z. (2022). Eelbrain: A python toolkit for time-continuous analysis with temporal response functions. *bioRxiv*.

Bryant, L., Ferguson, A., Valentine, M., and Spencer, E. (2019). Implementation of discourse analysis in aphasia: investigating the feasibility of a knowledge-to-action intervention. *Aphasiology*, 33(1):31–57.

Cocquyt, E. M., Vandewiele, M., Bonnarens, C., Santens, P., and De Letter, M. (2020). The sensitivity of event-related potentials/fields to logopedic interventions in patients with stroke-related aphasia. *Acta neurologica Belgica*, 120(4):805–817.

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., and Lalor, E. C. (2021). Linear modeling of neurophysiological responses to speech and other continuous stimuli: Methodological considerations for applied research. *Frontiers in Neuroscience*, 15.

Dalton, S. G., Stark, B. C., Fromm, D., Apple, K., MacWhinney, B., Rensch, A., and Rowedder, M. (2022). Validation of an automated procedure for calculating core lexicon from transcripts. *Journal of Speech, Language, and Hearing Research*, 65(8):2996–3003.

De Clercq, P., Vanthornhout, J., Vandermosten, M., and Francart, T. (2023). Beyond linear neural envelope tracking: a mutual information approach. *Journal of Neural Engineering*, 20(2):026007.

de Renzi, E. and Ferrai, C. (1978). The reporter's test: A sensitive test to detect expressive disturbances in aphasics. *Cortex*, 14(2):279–293.

Decruy, L., Vanthornhout, J., and Francart, T. (2019). Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. *Journal of Neurophysiology*, 122(2):601–615.

Desai, M., Field, A. M., and Hamilton, L. S. (2023). Dataset size considerations for robust acoustic and phonetic speech encoding models in eeg. *Frontiers in Human Neuroscience*, 16.

Devanga, S. R., Pollens, R. D., and Glista, S. O. (2021). Toward developing outcome measures in university-based aphasia programs: Perspectives from the aphasia communication enhancement program. *Perspectives of the ASHA Special Interest Groups*, 6(5):1047–1059.

Di Liberto, G. M. and Lalor, E. C. (2017). Indexing cortical entrainment to natural speech at the phonemic level: Methodological considerations for applied research. *Hearing Research*, 348:70–77.

Di Liberto, G. M., O'Sullivan, J. A., and Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, 25(19):2457–2465.

Dial, H. R., Gnanateja, G. N., Tessmer, R. S., Gorno-Tempini, M. L., Chandrasekaran, B., and Henry, M. L. (2021). Cortical tracking of the speech envelope in logopenic variant primary progressive aphasia. *Frontiers in Human Neuroscience*, 14.

Diedenhofen, B. and Musch, J. (2015). cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS ONE*, 10(3):e0121945.

Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.*, 19:158–164.

Ding, N. and Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *Journal of Neuroscience*, 33(13):5728–5735.

El Hachioui, H., Visch-Brink, E. G., de Lau, L. M., van de Sandt-Koenderman, M. W., Nouwens, F., Koudstaal, P. J., and Dippel, D. W. (2017). Screening tests for aphasia in patients with stroke: a systematic review. *Journal of Neurology*, 264(2):211–220.

El Hachioui, H., Visch-Brink, E. G., Lingsma, H. F., Van De Sandt-Koenderman, M. W., Dippel, D. W., Koudstaal, P. J., and Middelkoop, H. A. (2014). Nonlinguistic cognitive impairment in poststroke aphasia: A prospective study. *Neurorehabilitation and Neural Repair*, 28(3):273–281.

Etard, O. and Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *Journal of Neuroscience*, 39(29):5750–5759.

Fonseca, J., Raposo, A., and Martins, I. P. (2019). Cognitive functioning in chronic post-stroke aphasia. *Applied Neuropsychology: Adult*, 26(4):355–364.

Francart, T., Wieringen, A. V., and Wouters, J. (2008). Apex 3: a multi-purpose test platform for auditory psychophysical experiments. *J Neurosci Methods*, 172:283–293.

Fujioka, T., Ross, B., and Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *Journal of Neuroscience*, 35(45):15187–15198.

Gillis, M., Van Canneyt, J., Francart, T., and Vanthornhout, J. (2022). Neural tracking as a diagnostic tool to assess the auditory pathway. *Hearing Research*, page 108607.

Gillis, M., Vanthornhout, J., Simon, J. Z., Francart, T., and Brodbeck, C. (2021). Neural markers of speech comprehension: Measuring eeg tracking of linguistic speech representations, controlling the speech acoustics. *The Journal of Neuroscience*, 41(50):10316–10329.

Gillis, Kries, Vandermosten, M., and Francart, T. (2023). Neural tracking of linguistic and acoustic speech representations decreases with advancing age. *NeuroImage*, 267(119841):1–16.

Giraud, A. and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.*, 15:511–517.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., and Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, 11(12):e1001752.

Hamilton, L. S. and Huth, A. G. (2018). The revolution will not be controlled: natural stimuli in speech neuroscience. *Language, cognition and neuroscience*, 35(5):573–582.

Hyafil, A., Giraud, A.-L., Fontolan, L., and Gutkin, B. (2015). Neural cross-frequency coupling: Connecting architectures, mechanisms, and functions. *Trends in Neurosciences*, 38(11):725–740.

Ilvonen, T. M., Kujala, T., Tervaniemi, M., Salonen, O., Näätänen, R., and Pekkonen, E. (2001). The processing of sound duration after left hemisphere stroke: Event-related potential and behavioral evidence. *Psychophysiology*, 38:622–628.

Ince, R. A. A., Giordano, B., Kayser, C., Rousselet, G., Gross, J., and Schyns, P. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human brain mapping*, 38(3):1541–1573.

Jamal, N., Shanta, S., Mahmud, F., and Sha'abani, M. (2017). Automatic speech recognition (asr) based approach for speech therapy of aphasic patients: A review. *AIP Conference Proceedings*, 1883(1):020028.

Jansen, S., Luts, H., Wagener, K. C., Kollmeier, B., Del Rio, M., Dauman, R., James, C., Fraysse, B., Vormès, E., Frachet, B., Wouters, J., and van Wieringen, A. (2012). Comparison of three types of french speech-in-noise tests: A multi-center study. *International Journal of Audiology*, 51(3):164–173.

Johnson, L., Basilakos, A., Yourganov, G., Cai, B., Bonilha, L., Rorden, C., and Fridriksson, J. (2019). Progression of aphasia severity in the chronic stages of stroke. *American Journal of Speech-Language Pathology*, 28(2):639–649.

Kandylaki, K. D. and Bornkessel-Schlesewsky, I. (2019). From story comprehension to the neurobiology of language. *Language, Cognition and Neuroscience*, 34(4):405–410.

Kaufeld, G., Bosker, H. R., Alday, P. M., Meyer, A. S., and Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience*, 40(49):9467–9475.

Keitel, A., Gross, J., and Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology*, 16(3):e2004473.

Kertesz, A. (1982). Western aphasia battery. *New York: Grune and Stratton*.

Kim, H., Berube, S., and Hillis, A. E. (2022). Core lexicon in aphasia: A longitudinal study. *Aphasiology*, 0(0):1–13.

Kries, J., De Clercq, P., Lemmens, R., Francart, T., and Vandermosten, M. (2022). Tuning in on auditory details is difficult: Individuals with aphasia show impaired acoustic and phonemic processing. *bioRxiv*.

Kulasingham, J. P., Brodbeck, C., Presacco, A., Kuchinsky, S. E., Anderson, S., and Simon, J. Z. (2020). High gamma cortical processing of continuous speech in younger and older listeners. *NeuroImage*, 222:117291.

Lalor, E. C., Power, A. J., Reilly, R. B., and Foxe, J. J. (2009). Resolving precise temporal processing properties of the auditory system using continuous stimuli. *Journal of Neurophysiology*, 102(1):349–359.

Le, D., Licata, K., and Mower Provost, E. (2018). Automatic quantitative analysis of spontaneous aphasic speech. *Speech Communication*, 100:1–12.

Lesenfants, D., Vanthornhout, J., Verschueren, E., Decruy, L., and Francart, T. (2019). Predicting individual speech intelligibility from the neural tracking of acoustic- and phonetic-level speech representations. *Hearing Research*, 380:1–9.

Lesser, R. and Algar, L. (1995). Towards combining the cognitive neuropsychological and the pragmatic in aphasia therapy. *Neuropsychological Rehabilitation*, 5(1-2):67–92.

Lizarazu, M., Lallier, M., Bourguignon, M., Carreiras, M., and Molinaro, N. (2021). Impaired neural response to speech edges in dyslexia. *Cortex*, 135:207–218.

Lizarazu, M., Lallier, M., and Molinaro, N. (2019). Phase amplitude coupling between theta and gamma oscillations adapts to speech rate. *Annals of the New York Academy of Sciences*, 1453(1):140–152.

Mandke, K., Flanagan, S., Macfarlane, A., Gabrielczyk, F., Wilson, A., Gross, J., and Goswami, U. (2022). Neural sampling of the speech signal at different timescales by children with dyslexia. *NeuroImage*, 253:119077.

690 Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of eeg- and meg-data. *Journal of*
691 *Neuroscience Methods*, 164(1):177–190.

692 Mesik, J. and Wojtczak, M. (2022). The effects of data quantity on performance of temporal response
693 function analyses of natural speech processing. *bioRxiv*.

694 Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Jung, T. P., and Cauwen-
695 berghs, G. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry eeg. *IEEE*
696 *Transactions on Biomedical Engineering*, 11(62):2553–2567.

697 Ofek, E., Purdy, S. C., Ali, G., Webster, T., Gharahdaghi, N., and McCann, C. M. (2013). Processing of
698 emotional words after stroke: An electrophysiological study. *Clinical Neurophysiology*, 124:1771–1778.

699 Papathanasiou, I. and Coppens, P. (2017). Aphasia and related neurogenic communication disorders.
700 *Burlington, MA: Jones and Bartlett Learning*.

701 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Pretten-
702 hofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot,
703 M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*,
704 12:2825–2830.

705 Pedroni, A., Bahreini, A., and Langer, N. (2019). Automagic: Standardized preprocessing of big eeg
706 data. *NeuroImage*, 200(null):460–473.

707 Peelle, J. E. and Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension.
708 *Front Psychol*, 3.

709 Perez, A., Davis, M. H., Ince, R. A. A., Zhang, H., Zhanao, F., Lamarca, M., Lambon Ralph, M. A., and
710 Monahan, P. J. (2022). Timing of brain entrainment to the speech envelope during speaking, listening
711 and self-listening. *Cognition*, 224:105051.

712 Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). Iclabel: An automated electroencephalo-
713 graphic independent component classifier, dataset, and website. *NeuroImage*, 198(null):181–197.

714 Pulvermüller, F., Mohr, B., and Lutzenberger, W. (2004). Neurophysiological correlates of word and
715 pseudo-word processing in well-recovered aphasics and patients with right hemispheric stroke. *Psy-*
716 *chophysiology*, 41:584–591.

717 Robson, H., Pilkington, E., Evans, L., DeLuca, V., and Keidel, J. L. (2017). Phonological and semantic
718 processing during comprehension in Wernicke's aphasia: An N400 and Phonological Mapping Negativ-
719 ity Study. *Neuropsychologia*, 100(October 2016):144–154.

Rohde, A., Worrall, L., Godecke, E., O'Halloran, R., Farrell, A., and Massey, M. (2018). Diagnosis of aphasia in stroke populations: A systematic review of language tests. *PLoS ONE*, 13(3):e0194143.

Satopaa, V., Albrecht, J., Irwin, D., and Raghavan, B. (2011). Finding a "kneedle" in a haystack: Detecting knee points in system behavior. *2011 31st International Conference on Distributed Computing Systems Workshops*, pages 166–171.

Shannon, R. V., Zeng, F. G., V, K., Wygonski, J., and S, E. M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234):303–304.

Stark, B. C., Dutta, M., Murray, L. L., Fromm, D., Bryant, L., Harmon, T. G., Ramage, A. E., and Roberts, A. C. (2021). Spoken discourse assessment and analysis in aphasia: An international survey of current practices. *Journal of Speech, Language, and Hearing Research*, 64(11):4366–4389.

Søndergaard, P., Torrésani, B., and Balazs, P. (2012). The linear time frequency analysis toolbox. *International Journal of Wavelets Multiresolution and Information Processing*, 10:1250032.

Van Ewijk, E., Dijkhuis, L., Hofs-Van Kats, M., Hendrickx-Jessurun, M., Wijngaarden, M., and De Hilster, C. (2020). *Nederlandse Benoem Test*. Bohn stafleu van loghum.

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., and Francart, T. (2018). Speech intelligibility predicted from neural entrainment of the speech envelope. *JARO - Journal of the Association for Research in Otolaryngology*, 19(2):181–191.

Verschueren, E., Gillis, M., Decruy, L., Vanthornhout, J., and Francart, T. (2022). Speech understanding oppositely affects acoustic and linguistic neural tracking in a speech rate manipulation paradigm. *Journal of Neuroscience*, 42(39):7442–7453.

Visch-Brink, E., Van de Sandt-Koenderman, M., and El Hachioui, H. (2010). *ScreeLing*. Houten: Bohn Stafleu Van Loghum.

Wallace, S. E. and Kimelman, M. D. (2013). Generalization of word retrieval following semantic feature treatment. *Neurorehabilitation*, 32(4):899–913.

Wöstmann, M., Lim, S., and Obleser, J. (2017). The human neural alpha response to speech is a proxy of attentional control. *Cerebral cortex*, 27(6):3307–3317.

Xu, N., Zhao, B., Luo, L., Zhang, K., Shao, X., Luan, G., Wang, Q., Hu, W., and Wang, Q. (2022). Two stages of speech envelope tracking in human auditory cortex modulated by speech intelligibility. *Cerebral Cortex*, 33(5):2215–2228.

Zan, P., Presacco, A., Anderson, S., and Simon, J. Z. (2020). Exaggerated cortical representation of speech in older listeners: mutual information analysis. *Journal of Neurophysiology*, 124(4):1152–1164.

751 Zou, J., Xu, C., Luo, C., Jin, P., Gao, J., Li, J., Gao, J., Ding, N., and Luo, B. (2021). ϑ-band cortical
752 tracking of the speech envelope shows the linear phase property. *eNeuro*, 8(4).