The spatio-temporal dynamics of phoneme encoding in aging and aphasia

Jill Kries[1], Pieter De Clercq[2], Maaike Vandermosten[2]* and Laura Gwilliams[1]*

Affiliations:

[1] Department of Psychology, Stanford University, Stanford, CA, USA

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

* shared last author

Corresponding author: Jill Kries, jill.kries@stanford.edu

Email addresses: jill.kries@stanford.edu; pieter.declercq@kuleuven.be; maaike.vandermosten@kuleuven.be; laura.gwilliams@stanford.edu

Acknowledgements: The authors would like to thank all participants, especially all the brave participants with aphasia and their partners, family or friends that support them. Moreover, the authors would like to thank everyone who helped with data collection and recruitment: Dr. Klara Schevenels, Janne Segers, Rosanne Partoens, Charlotte Rommel, Dr. Ramtin Mehraram, Ines Robberechts, Laura Van Den Bergh, Anke Heremans, Frauke De Vis, Mouna Vanlommel, Naomi Pollet, Kaat Schroeven, Pia Reynaert, and Merel Dillen.

Funding Sources: Pieter De Clercq was financially supported by the Research Foundation Flanders (FWO; PhD grant: SB 1S40122N); Maaike Vandermosten was financially supported by the FWO grant No. G0D8520N; Laura Gwilliams was supported by BRAIN Foundation A-0741551370 and Whitehall Foundation 2024-08-043.

Conflict of interest statement: No conflicts of interest, financial or otherwise, are declared by the authors.

Ethics approval and patient consent statement: The study received ethical approval by the medical ethical committee of KU Leuven and UZ Leuven (S60007) and is in accordance with the declaration of Helsinki. Informed consent was obtained from all participants for the recruitment via screening and for the data collection in the chronic phase post-stroke.

Author contribution statement: Conceptualization: JK, LG; Data Acquisition: JK, PDC, MV; Project Administration: JK; Data Curation: JK; Resources: MV, LG; Methodology: LG, JK; Formal analysis: JK, LG; Visualization: JK; Writing – Original Draft Preparation: JK; Writing – Review and Editing: JK, PDC, MV, LG; Supervision: LG; Funding Acquisition: MV, LG.

Abstract

During successful language comprehension, speech sounds (phonemes) are encoded within a series of neural patterns that evolve over time. Here we tested whether these neural dynamics of speech encoding are altered for individuals with a language disorder. We recorded EEG responses from individuals with post-stroke aphasia and healthy age-matched controls (i.e., older adults) during 25 min of natural story listening. We estimated the duration of phonetic feature encoding, speed of evolution across neural populations, and the spatial location of encoding over EEG sensors. First, we establish that phonetic features are robustly encoded in EEG responses of healthy older adults. Second, when comparing individuals with aphasia to healthy controls, we find significantly decreased phonetic encoding in the aphasic group after shared initial processing pattern (0.08-0.25s after phoneme onset). Phonetic features were less strongly encoded over left-lateralized electrodes in the aphasia group compared to controls, with no difference in speed of neural pattern evolution. Finally, we observed that phonemes with high uncertainty about word identity were encoded longer in controls than in individuals with aphasia. This indicates that encoding phonetic information until word identity is resolved might be a crucial mechanism for successful speech comprehension. Together, our results suggest that aphasia may entail failure to maintain lower-order information long enough to recognize lexical items.

Significance statement

This study reveals robust phonetic decoding with EEG recordings in older adults, as well as decreased phonetic processing in individuals with a language disorder (aphasia) compared to healthy controls. This was most prominent over left-sided electrodes. Additionally, we observed that healthy controls, but not individuals with aphasia, encode phonetic features longer when uncertainty about word identity is high, indicating that this mechanism - encoding phonetic information until word identity is resolved - is crucial for successful language processing. These insights deepen our understanding of disrupted mechanisms in a language disorder, and show how the integration between language processing levels works in the healthy aging, neurotypical brain.

Keywords: phoneme decoding, speech processing, aphasia, EEG, natural language processing

1 Introduction

Comprehending natural speech entails rapidly converting continuous acoustic input into discrete units, such as phonemes and words (Hickok and Poeppel, 2007). A study by Gwilliams et al. (2022) demonstrated that the spectrotemporal properties of phonemes, known as phonetic features, are dynamically encoded in the brain. This coding scheme facilitates integration with higher-level language represen-

tations, such as syntactic or semantic features, through hierarchical dynamic coding (Gwilliams et al., 2024). Individuals with language disorders such as aphasia can have deficits in phoneme identification (Kries et al., 2023) or phonological processing (Blumstein, 1998), among other difficulties. With the hierarchical dynamic coding framework, we now have a unique opportunity to directly investigate the neural dynamics of speech representations in individuals with aphasia. This paper aims to address this research gap by investigating the effects of aphasia on phonetic feature encoding.

When healthy young adults listen to speech, phonetic features are encoded in auditory cortex, including the transverse temporal gyrus, superior temporal gyrus, and superior temporal sulcus (Mesgarani et al., 2007, 2014; Wilson et al., 2018; Levy and Wilson, 2019; Yi et al., 2019). Based on magnetoencephalography (MEG) data during natural story listening, Gwilliams et al. (2022) showed that the spectro-temporal properties of the phonetic features are encoded for around 0.3 s, outlasting the duration of the phoneme input, and that the encoded information is passed to a new neural ensemble every 0.08 s, which is proportional to the average phoneme duration in the stimulus (Gwilliams et al., 2022, 2024). The authors conclude that phonetic features are encoded in a dynamic neural pattern, whereby different neural ensembles are recruited across time – referred to as hierarchical dynamic coding (HDC) (Gwilliams et al., 2024). This means that rather than a one-to-one mapping between neural ensemble and feature encoding, different ensembles are activated in sequence. This allows a running history of the past three phonemes to be encoded in parallel, while also keeping track of the relative order that they were said.

Aphasia is often considered primarily a phonological, semantic or syntactic language disorder, due to commonly observed symptoms like word finding difficulties or incomplete sentences (Harnish, 2018; Zhang and Hinzen, 2022). However, recent work suggests that the processing issues associated with aphasia can also be related to more low-level speech processing, i.e., neural encoding of amplitude fluctuations of speech and phoneme identification (Kries et al., 2023, 2024). This implies that some language difficulties experienced in aphasia may arise from an impoverished representation of speech sounds. In fact, Kries et al. (2023) found that individuals with aphasia (IWA)'s performance at discriminating between different rates of change in amplitude predicted phonological processing performance, suggesting that low-level processing impairments may propagate onto higher-level processes. Building upon this prior work, we aim to assess whether difficulties in phoneme processing in aphasia may stem from challenges in dynamic coding of phonetic features.

Gwilliams et al. (2022) used MEG recordings to investigate phonetic encoding during a two-hour story-listening task in younger adults. By contrast, we are using an available electroencephalography (EEG) dataset of individuals with aphasia and healthy age-matched controls listening to natural speech for 25 minutes (fig.1B) (Kries et al., 2024). EEG is more feasible than MEG for post-stroke or older adults, because it is more accessible in clinical settings, and allows for more flexible body position during the

recording. Moreover, shorter neural recordings are desirable to maintain optimal attention levels. Therefore, our first research question examines whether the robust phonetic decoding observed by Gwilliams et al. (2022) in younger adults can be replicated with (1) EEG data, (2) a shorter recording duration and (3) in older adults.

The second research question examines the dynamics of phonetic encoding in IWA compared to age-matched healthy controls. Specifically, we wanted to know whether IWA would show a change in phonetic encoding duration, a change in speed of neural pattern evolution, or a mix of both (fig.1F). We hypothesized slower speed of neural pattern evolution in aphasia, because that would result in representational overlap between neighbouring phonemes and might explain symptoms observed in aphasia, such as phonemic paraphasias. Given the structural and functional neural changes following stroke-induced aphasia (Wilson et al., 2023), we also test for differences in topography of neural activity during phonetic encoding between IWA and healthy controls.

2 Results

To answer these research questions, 39 IWA in the chronic phase after stroke (> 6 months after onset; age in years: mean=69.5±12.4) and 24 age-matched healthy control participants (age in years: mean=71.5±7) listened to a narrative for 25 minutes (in their native language, Dutch) while EEG data was recorded (fig.1B). Figure 1A depicts the performance on the Dutch picture-naming test by group (Van Ewijk et al., 2020). The narrative, which comprised 13,560 phonemes and 3,002 words, was annotated for phonetic features, allowing us to apply time-resolved decoding of phonetic features from the EEG signals (fig.1C, D). All decoding was performed using one-versus-all logistic regression and 5-fold cross-validation, using a temporal generalization approach, i.e., evaluate the classifier's accuracy across all possible train/test time combinations (fig.1E). We apply this to 18 phonetic features, across manner of articulation, place of articulation, voicing, roundness and front-backness of consonants and vowels (fig.1C; section 5).

A note on using the terms decoding/encoding: Here we are applying a decoding analysis, i.e., the input to the regression models are the brain data recorded from EEG sensors, and the output of the models are the speech features. We would like to underline that when a feature can be decoded from brain data, that commensurately entails that the feature is *encoded* in the neural activity. We choose a decoding regime because (i) part of our aim is to replicate the previous study that also used this regime; (ii) leveraging multivariate EEG data provides robustness against noise and aggregation of shared and complimentary signal across channels (King et al., 2020).

Throughout the paper, we are referring to healthy control participants and older adults interchangeably given that these participants are on average 71 years old, with a range of 58 to 87 years of age.

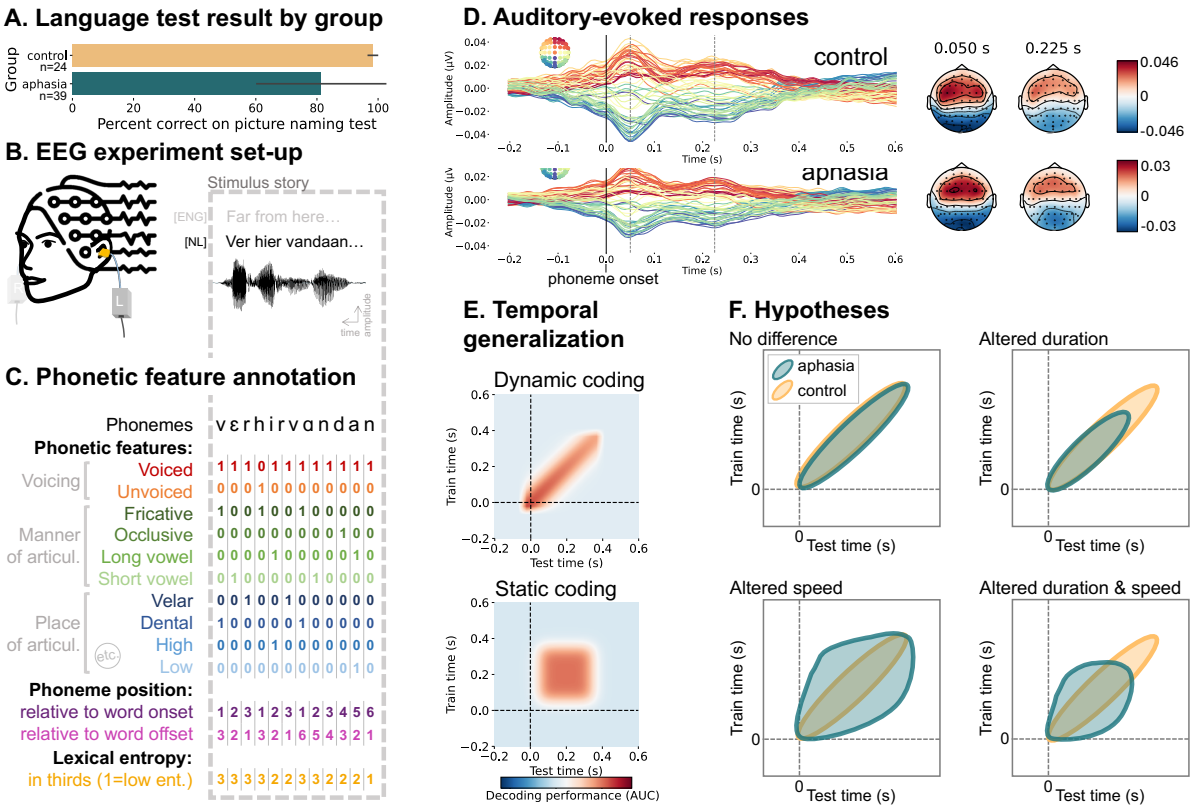

Figure 1: **A.** Mean and standard deviation for the picture naming test scores (in percentage) by group. **B.** The experimental protocol consisted of participants listening to a 25 minute-long narrative in Flemish while 64-sensor-EEG data was being recorded. The audio was presented bilaterally. The first 3 words and their acoustic waveform are depicted, with an English translation. **C.** The narrative was annotated for 18 phonetic features (not all illustrated) with across-feature time alignment. Moreover, the phonemes were subset by position within word and by lexical entropy (entropy values were categorized into thirds). **D.** The EEG data was epoched around the phoneme onsets. This plot displays the time courses of the auditory evoked response for all 64 EEG sensors on the left and the topographies at 0.05s and 0.225s after phoneme onset on the right, for the control group and aphasia group separately. **E.** Temporal generalization: Training a decoder at each time slice and testing each of these decoders across all possible time slices allows to investigate the encoding duration (matrix diagonal) and generalization (width of significant cluster). The top plot shows a dynamic encoding pattern wherein the encoded information evolves across temporal decoders, i.e., the way that information is encoded changes over time. The bottom plot shows a static encoding pattern wherein the encoded information is maintained. **F.** We hypothesized that the aphasia group would show a different encoding pattern than the healthy control group. The top right graph illustrates what the temporal generalization (TG) pattern would look like if IWA had a shorter encoding duration than controls. The bottom left graph shows what we would expect to see if IWA had a slower speed of evolution of encoding than controls, i.e., longer generalization. The bottom right graph displays what we would expect to observe if both duration and generalization of encoding would be altered in aphasia.

2.1 Replicating phonetic encoding dynamics in older adults with EEG data

Gwilliams et al. (2022) found that phonetic features are represented in neural activity of younger adults, measured with MEG data during 2 hours of story listening. Specifically, the time window in which phonetic features were significantly decodable from the MEG data was from 0.05 to 0.3 s after phoneme onset. To test whether phonetic features can be robustly decoded in healthy older adults using EEG data during 25 minutes of story-listening, we epoched the EEG data around each phoneme (-0.2 to 0.6 s around onset). We decoded phonetic features from the EEG data, specifically for every epoch time point independently (see Methods for more details on the implementation of the decoding analysis). This resulted in a time course of decoding performance, reflecting to what extent the phonetic information is present in the EEG signal, for each of the 18 tested phonetic features. We assessed significance of the decoding time course of phonetic features using one-sample permutation cluster tests across subjects. We found that all 18 phonetic features were decodable above chance (cluster-forming threshold: p=.05; results in table 1 and fig.2A). Due to variations in phoneme duration, such as between vowels and consonants, and the differing number of phoneme trials within certain phonetic feature categories, there is variability in decoding performance between features. Therefore, we focus on the decoding performance averaged across all features moving forward. When averaging across phonetic features, we observed above chance decoding from -0.04 to 0.49 s ($p<.001$) relative to phoneme onset for the control group (fig.2B top right plot). Thus, on average, phonetic features are decodable for a span of 0.53 s. Given that the average phoneme duration in the story was 0.098 s (SD=0.107 s), this replicates the prior finding from younger adults (Gwilliams et al., 2022) that the neural encoding of phonetic features long surpasses the duration of the phoneme itself. Results hereafter are all based on the average across phonetic features.

The implication of sustained neural encoding is that multiple phonemes are processed simultaneously. Gwilliams et al. (2022) found that at any moment in time, the brain encodes the phonetic information of the preceding 3 phonemes, and the order of phonemes is retained by jointly encoding the surpassed time passed since onset. Gwilliams et al. (2022) were able to prove this by employing TG analysis, enabling insight into how fast neural configurations that encode information change. Specifically, a given time slice (of the epoched EEG data) used for *training* the classifier corresponds to one specific neural configuration, and *testing* the given classifier across all time slices tells us for how long the phonetic information is maintained within the same neural configuration - and in turn the speed at which the information evolves. Gwilliams et al. (2022) found that phonetic information was maintained within a same neural configuration for a duration of 0.08 s, and thus a new cycle of neural evolution starts every 0.08 s. This demonstrates that the neural representation of phonetics is evolving dynamically (fig1E). Here, we wanted to investigate whether healthy older adults would also show a dynamic processing scheme, given that this seems to be an important mechanism for orderly processing of phoneme sequences. Moreover, this coding principle for phonetics has not yet been investigated with EEG.

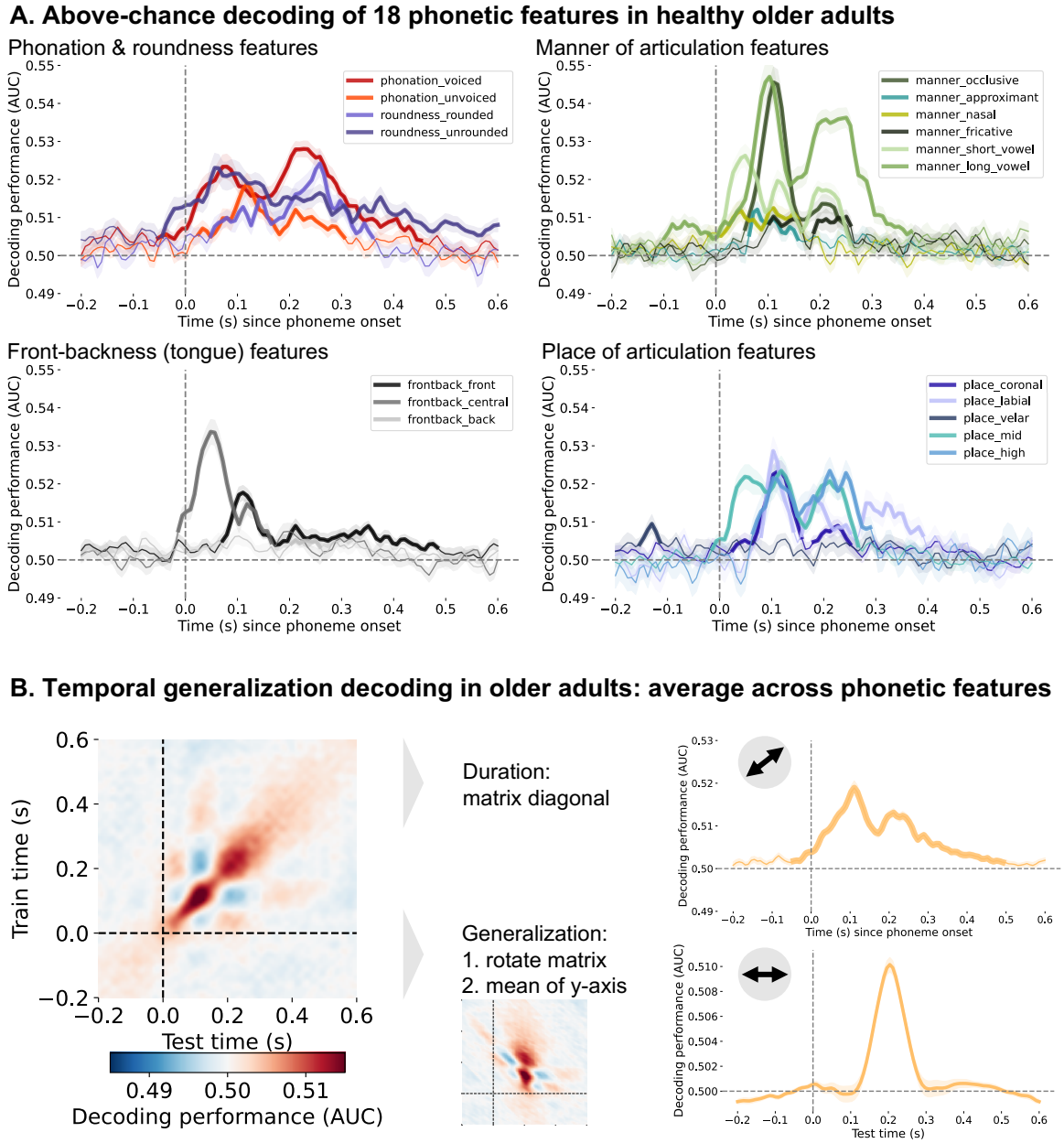

Figure 2: **A.** Above-chance decoding of 18 phonetic features in healthy older adults. The features vary in decoding performance and over time. The thicker line segments indicate the time points where the features are significantly decoded above chance. **B.** The TG approach in healthy older adults revealed that phonetic features are decodable above chance for 0.5s (top right panel) and generalizable for 0.139s (bottom right panel). The decoding duration corresponds to the diagonal of the TG matrix (left panel), thus the decoding performance when the decoder was trained on the same time slice as was used for testing the decoder. To quantify the generalization time, the TG matrix is reoriented and then averaged across the y-axis.

phonetic feature	significant time window (s)	average decoding performance (AUC)	average t-value	p-value
voiced	-0.052 to 0.452	0.515	1.414	<.001
unvoiced	0.025 to 0.304	0.509	-0.009	<.001
nasal	-0.013 to 0.165	0.508	0.785	<.001
fricative	0.102 to 0.165	0.508	-1.418	0.003
	0.188 to 0.258	0.508	0.497	<.001
occlusive	0.056 to 0.258	0.518	1.662	<.001
approximant	0.064 to 0.157	0.508	2.42	<.001
short vowel	-0.013 to 0.258	0.514	1.029	<.001
long vowel	-0.099 to 0.374	0.518	1.258	<.001
coronal	0.025 to 0.157	0.511	1.905	<.001
	0.180 to 0.250	0.506	1.220	<.001
labial	0.079 to 0.141	0.516	0.758	0.002
	0.157 to 0.250	0.510	-0.077	0.002
	0.273 to 0.390	0.511	-0.963	<.001
velar	-0.153 to -0.114	0.506	1.059	0.009
mid	-0.013 to 0.266	0.515	-0.484	<.001
high	0.064 to 0.289	0.516	-1.238	<.001
rounded	0.048 to 0.359	0.511	-0.196	<.001
unrounded	-0.044 to 0.6	0.512	2.255	<.001
front	0.071 to 0.483	0.506	1.406	<.001
central	-0.013 to 0.165	0.517	0.686	<.001
back	0.203 to 0.258	0.504	1.307	0.003

Table 1: Above-chance decoding of phonetic features in healthy older adults

For the TG analysis, we decoded phonetic features from the epoched EEG data, and trained the classifier at a given time point and tested it on all time points in the epoch. We then repeated this process for all train times, thus decoding phonetic features at each possible combination of train and test time points. This results in a train time by test time matrix wherein each data point represents the decoding performance (fig.2B left panel). The width of the significant cluster in this matrix represents the generalization time of the phonetic information in a given time point and spatial configuration (i.e., a decoder trained at time t). To extract the duration of generalization, we rolled the rows of the matrix such that the values along the diagonal become oriented at a given x coordinate. We then average across the y-axis of the matrix, resulting in fig.2B right bottom panel. The width of the significant matrix cluster corresponded to 0.139 s in healthy older adults. Same as in Gwilliams et al. (2022), we compared the phonetic encoding duration (0.53 s, the matrix diagonal) to the generalization (0.147 s, the width of the reoriented matrix using a one-sample permutation cluster test to confirm the dynamic evolution of phonetic encoding. We found a significant difference (cluster 1: t=-2.86, p<.001, time: -0.044s to 0.165s; cluster 2: t=-3.12, p<.001, time: 0.188s to 0.46s), confirming that phonetic information is encoded dynamically (fig.1E top panel), evolving over time and space, in healthy older adults.

The same analysis and figures for the aphasia group can be found in the supplementary information (fig.S.2).

2.2 Phonetic encoding in individuals with aphasia

IWA have been found to have impaired phoneme identification (Kries et al., 2023) and phonological processing (Blumstein, 1998). This led us to investigate the neural dynamics of phonetic encoding in aphasia in comparison to healthy controls. Regarding the TG analysis, we expected that the aphasia group would show a different encoding pattern than the healthy control group (fig.1F). We tested whether IWA would show shorter encoding duration, slower speed of evolution of encoding than controls, or a mix of both. Figure 3A shows the TG matrices for the aphasia and control groups overlaid (depicted as contours around areas with a high decoding performance). We conducted a permutation cluster test to compare matrices between groups and found a cluster in which the aphasia group showed significantly lower decoding performance than the control group ($p<.001$).

The diagonal of the matrix, i.e., where the classifier is trained and tested on the same epoch time point, indicates how long phonetic features are decodable. Thus, we tested whether IWA show an altered duration of decodability. We compared the aphasia group to the control group using a permutation cluster test, and found significantly lower decoding accuracy in the aphasia group compared to the control group between 0.08 and 0.25 s relative the phoneme onset ($p<.001$; fig.3B). Thus, the aphasia group shows lower encoding at relevant processing time points (also see supplementary fig.S.3). For a between-group comparison of above-chance decodability duration, see supplementary section S.5. Based on the two observed peaks in the decoding time course, we additionally investigated whether the decoding performance at each peak correlates with different behavioral measures, within the aphasia group. We hypothesized that behavioral correlates of each peak could provide insight into the associated functional processes; however, no significant correlations were found between either peak and language- or cognition-related variables (supplementary fig.S.4).

To compare groups on how long phonetic information is maintained within a same temporal decoder and spatial configuration, we extracted the train time lags between 0 and 0.35 s, based on previous research (Gwilliams et al., 2022), and standardized the data by subtracting the mean and dividing by the standard deviation (using StandardScaler from scikit-learn; fig.3A). We used a permutation cluster test to compare groups on generalization time, and found no significant difference between groups (fig.3C), with average generalization time being 0.147 s in the aphasia group and 0.139 s in the control group. The speed of evolution of phonetic encoding across spatial configurations in IWA is highly overlapping with healthy controls. Taken together, these results partially support the hypothesis in the top right panel of figure 1F: an altered duration of phonetic encoding in aphasia (i.e., weaker encoding after initial processing time), but no change to the speed with which the phonetic information is passed between neural populations. However, the difference in duration of phonetic encoding that we observed is indirect, i.e., phonetic information is more weakly encoded in aphasia after initial processing time, a scenario that we did not

fully capture in our hypotheses in figure 1F.

Finally, to examine the spatial underpinning of these dynamic representations, we analyzed the decoding performance at every sensor. We used the time course of activity as the input features to the logistic regression, fit for each EEG sensor separately. We submitted the decoding performance of each sensor across subjects and groups to a mass-univariate independent samples t-test (Brodbeck, 2020) to analyze topographical differences of phonetic feature decoding between the aphasia group and the control group (fig.3D). With a cluster-forming p-value threshold of p<.001, we identify a cluster of 22 EEG sensors (p<.001; fig.3D) over which IWA had lower decoding accuracy than healthy controls. The sensors primarily localized to left auditory cortex and posterior sensors bilaterally.

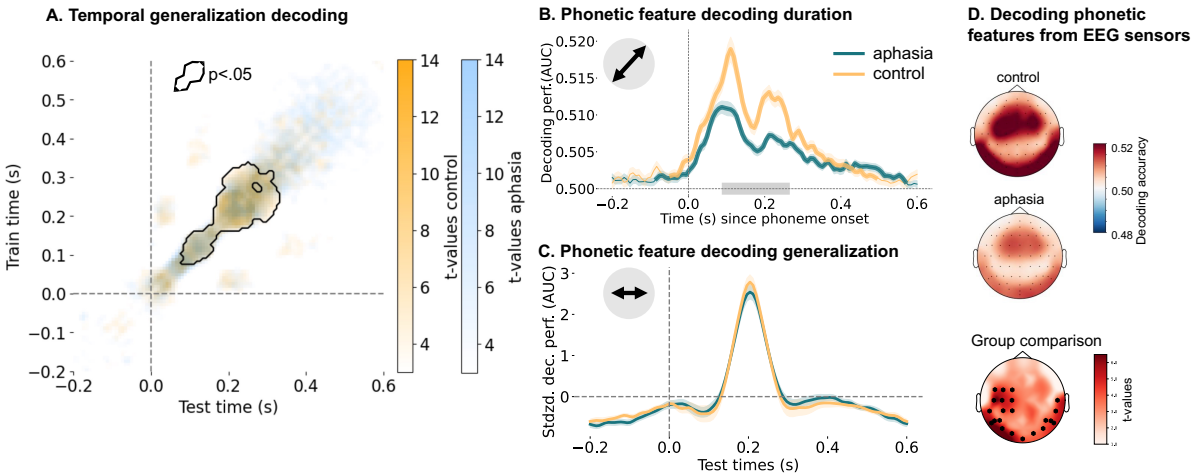

Figure 3: **A.** The TG matrices shown as t-values overlaid for the control (yellow) and aphasia (blue) group. To obtain this, a one sample permutation cluster test was conducted for each group separately, which calculates a t-value for each data point in the TG matrix. The black contour indicates the cluster in which the groups significantly differed from each other. To achieve this, we conducted a permutation cluster test to compare groups. **B.** A significant difference between groups was found from 0.08 to 0.25 s (as indicated by the gray bar) where the control group showed increased decoding performance. The bold parts of the lines indicate the time points where the features are significantly decoded above chance. **C.** No significant difference in generalization was found between groups. For both groups, the encoded information was passed on to new neural ensembles every 0.14 s. **D.** The top 2 topographies show how well phonetic features could be decoded for each of the 64 EEG sensors. The bottom topography shows the difference between both groups, specifically the t-values calculated using a mass-univariate independent samples t-test. The control group showed generally higher decoding, but especially in the 22 sensors marked in bold in the bottom topography (p<.001).

2.3 Effects of lexical predictability on phonetic encoding

Given that we observed a difference in the duration of phonetic encoding in IWA compared to healthy controls, we decided to investigate post hoc what might cause this alteration. Previous research has found that lexical predictability plays a role in the duration of phonetic encoding (Gwilliams et al., 2022). Certainty about the lexical outcome can be quantified by computing the Shannon entropy of the activated cohort of words given the preceding phoneme inputs, here referred to as lexical entropy (Gwilliams and Davis, 2022). Phonemes at the beginning of words generally have a higher degree of

entropy (uncertainty) about lexical outcome because there are more possible words in the activated cohort. By contrast, phonemes later in the word often entail less uncertainty about lexical identity because there are fewer possible outcomes.

Based on these previous results, we investigated whether changes in phonetic encoding duration in aphasia are related to an altered interaction between phonetic encoding and lexical uncertainty. If so, we should observed that encoding duration is contingent upon certainty about word identity – the more uncertainty there is about lexical identity, the longer the phonetic information is encoded (Gwilliams et al., 2022). We conducted two analyses to investigate this; (1) we subset all phonemes according to phoneme position relative to word onset (first phoneme within word, second one, etc.) and offset (last phoneme of the previous word, second to last one, etc.) and (2) we subset phonemes according to lexical entropy, namely comparing phonemes with the highest and lowest entropy values.

To assess phonetic encoding for each phoneme position relative to word boundaries, we fit the classifier using the full training dataset (80% of 13,560 phonemes). Because words are of varying lengths, this means that the test set gets increasingly smaller as the phoneme position moves away from word boundaries (see table 2 for specific numbers of trials included in each analysis). We compared groups on each phoneme position separately using a permutation cluster test, demonstrating a significant group difference for the first ($p<.001$), second ($p=.02$) and third ($p=.03$) phoneme positions, as well as the second to last ($p<.01$) and fifth to last ($p=.016$) phoneme positions (fig.4A). For the first phoneme relative to word onset, we observed a difference in reactivation, wherein only the control group seemed to maintain longer or "reactivate" the phonetic features in earlier temporal decoders, i.e., 0.1 to 0.25 s (supplementary fig.S.5). For the other phoneme positions, the group difference was located on or close to the TG matrix diagonal, thus displaying similar results as found when taking into account all phonemes, i.e., altered strength of encoding. We would like to point out that we might not have had enough phoneme trials beyond the second (and second to last) phoneme position (table 2) to draw strong conclusions from this analysis.

To investigate the effect of lexical entropy on phonetic encoding, we fit the classifier again using the full training dataset, while the test set consisted of phonemes with low (20% of 3,427 phonemes) or high entropy (20% of 3,592 phonemes). Entropy is one possible way to measure linguistic prediction (Gwilliams and Davis, 2022), which is represented within the language network (Shain et al., 2020), and given that the language network is damaged in aphasia, we expect the linguistic predictions mechanism to be disrupted. Therefore, we hypothesized that IWA would show no difference in phonetic decoding performance between phonemes with low and high entropy, but that healthy controls would show such a difference.

To compare groups and entropy conditions, we first compared groups for the low and high entropy conditions separately using a permutation cluster test. We found that for both entropy conditions, the

control group has higher decoding than the aphasia group between 0.1s and 0.15s after phoneme onset (low entropy: 0.110s-0.157s, average decoding: control=0.013 and aphasia=0.008, t=0.757, p=0.042; high entropy: 0.095s-0.149s, average decoding: control=0.009 and aphasia=0.004, t=0.002, p=0.011; fig.4B). Additionally, for phonemes with high entropy only, the control group shows also higher decoding in a window from 0.188s to 0.219s (average decoding: control=0.012 and aphasia=0.004, t=1.025, p=0.002) after phoneme onset (fig.4B). Second, we tested the interaction between group and entropy condition by subtracting the low entropy condition from the high entropy condition (see supplementary fig.S.7) for each group, subsequently conducting an independent samples t-test for each time lag and finally correcting p-values for multiple comparisons (false discovery rate). Based on findings from (Gwilliams et al., 2022), we tested in a time window from 0.150s to 0.350s. We found a significant interaction between group and entropy condition from 0.204s to 0.212s after phoneme onset (t=2.145, p=0.037), wherein the control group had higher decoding performance than the aphasia group in the high entropy condition only. This means that phonemes with higher lexical uncertainty are encoded for longer – an adaptive encoding duration mechanism found in previous work (Gwilliams et al., 2022). Here we replicate this finding in healthy old adults, but find no evidence for the same processing occurring for people with aphasia.

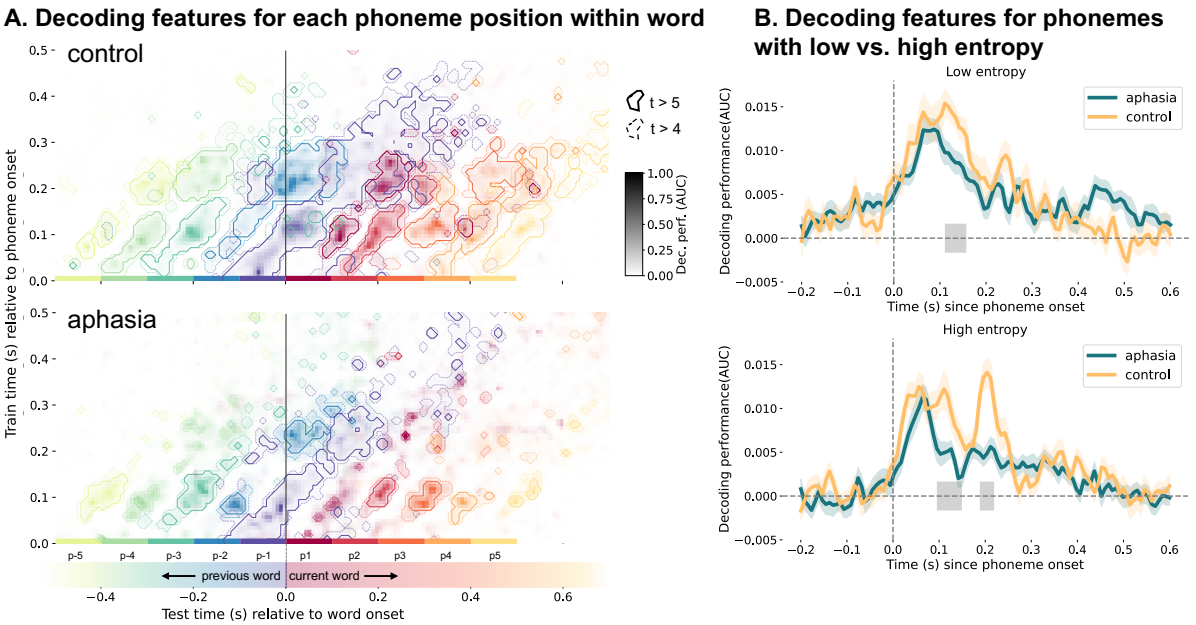

Figure 4: **A.** The top plot shows TG decoding of phonetic features for each phoneme position within word separately. Each phoneme position is depicted in a different color. The bottom plot shows the same type of plot for the aphasia group. The TG matrices are shown as contour plots, where the contours frame t-values larger than 5 (solid lines) and 4 (dashed lines). The decoding strength is shown as gradient from light (lower decoding) to dark (higher decoding) in the respective color of each phoneme position. Phoneme positions where the decoding matrices significantly differed between groups are P1, P2, P3, P-2 and P-5. **B.** We decoded phonetic features separately for phonemes with low and high lexical entropy. The top plot shows the difference between groups for low entropy phoneme decoding, the bottom plot shows the same graph for high entropy phoneme decoding. The gray bars indicate where the groups are significantly different from each other.

3 Discussion

In this study, we decoded phonetic features from EEG data during continuous story listening in 39 individuals with aphasia (IWA) and 24 age-matched (thus older) healthy controls. We found (1) phonetic features can be robustly decoded from 25 minutes of EEG data in healthy older adults, (2) IWA have an altered phonetic encoding duration, strength and topography compared to controls, and (3) the adaptive mechanism of phonetic encoding duration being contingent on lexical uncertainty seems to be disrupted in aphasia.

Robust encoding of phonetic features was reproduced in older adults using an EEG dataset (fig.2), which replicates a prior MEG study with younger adults (Gwilliams et al., 2022). This is an important result, given that the signal-to-noise ratio is higher for MEG than EEG data (Goldenholz et al., 2009; Ahlfors et al., 2010), and given the 3.7x difference in data size between studies, i.e., 2 hours and 50,518 phonemes in Gwilliams et al. (2022) versus 25 minutes and 13,560 phonemes in the present study. The fact that the study by Gwilliams et al. (2022) used English stimuli and here we used Dutch stimuli may also introduce variation in results due to differing phonetic features and their statistical probabilities and frequency. The different phonetic features show slightly distinct decoding time courses, e.g., short vowels have an earlier first peak than long vowels, and phonemes with rounded lips have a later decoding peak than phonemes with unrounded lips. All phonetic features were decodable above chance at some point throughout the time course. Finally, we also replicated the prior finding of dynamic coding of phonetic features, in older adults. We found that the average decoding across features was above chance for 0.53 s, whereas the duration with which the same neural pattern remained informative was only 0.139 s. This suggests that in both younger and older adults, a dynamic neural pattern is used to encode and process the phonetic features of speech sounds in continuous speech.

Comparing decoding duration between groups, we found that phonetic features were decoded less well in IWA than in controls at certain time points (fig.3B). Early during processing (0 to 0.08 s), no significant difference between groups was found (fig.3A&B); the difference emerged from 0.08-0.25 s. IWA may initially process phonetic features, but not keep the information encoded long enough to fully integrate it with higher-level processes (Gilbert and Sigman, 2007; Sigman and Dehaene, 2005; Gwilliams et al., 2018). Thus, the duration of encoding is indirectly affected in aphasia, via lower encoding at the later processing time points.

Comparing the TG decoding performances of both groups (fig. 3A and supplementary fig.S.3), it seems that the decoding performance is more diffuse, less concentrated in the aphasia than the control group. More diffuse cognitive representations in the brain have been found to be a part of the aging process, a mechanism called dedifferentiation (Koen and Rugg, 2019). Dedifferentiation refers to a loss of functional specialization in neural structures, meaning that networks that were once specialized in a specific cognitive

function become less specialized and show more diffuse neural activity in older age. Our control sample consisted of older adults (on average 71 years old), same as our aphasia sample, but IWA might show more dedifferentiation than controls. Could dedifferentiation also play a role in aphasia after stroke? Purcell et al. (2019) found that increased local neural response differentiation was related to less severe language impairments. This might explain the more smeared, less specific neural response to phonetic features in aphasia compared to healthy controls. Another possibility is that the EEG data in IWA is affected by the stroke lesion site being filled with cerebro-spinal fluid, which in turn impacts the signal-to-noise ratio over sensors closest to the lesion site (Piastra et al., 2022; Salinet et al., 2014; Zbesko et al., 2018; Vorwerk et al., 2014; Cohen et al., 2015; Park et al., 2016; Cassidy et al., 2020).

The sensor-level decoding analysis unveiled that, while similar sensors encode phonetic information across groups (fig.3D), there is stronger encoding in healthy controls, which is most pronounced over left auditory sensors. These results are in line with previous findings of decreased speech envelope encoding in the same aphasia dataset (Kries et al., 2024). While the stroke lesion in the left hemisphere (see supplementary fig.S.1 for lesion overlap map) might influence the decoding accuracy over those sensors, it is also possible that there is a functional reduction in processing speech features. With the current dataset and analysis, it is not possible to disentangle whether the effect stems from structural or functional changes, or a mix of both. A third control group consisting of stroke patients with a right-hemispheric lesion without aphasia could help to disambiguate whether the effects that we see are due to structural or functional changes in the brain.

Given that phonological impairments, such as mixing up sequence order or phoneme misidentification, have been found in IWA (Blumstein, 1998), we expected to see a more square-shaped TG matrix (fig.1E), meaning that IWA would have less dynamic encoding of phonetic features and pass on the information at a slower pace between neural populations, creating more representational overlap between phonemes of a sequence. However, we did not find evidence for more representational overlap between sequential phonemes. Instead, we observed that the neural pattern evolves at the same rate for both groups (fig.3). Specifically, the neural pattern completes a full evolutionary cycle every 0.14 s. This implies that phonological impairments in IWA are not related to not being able to keep apart the phonetic representations of neighbouring phonemes at a neural level.

Aphasia is a very heterogeneous disorder, with participants spanning from mild to severe aphasia and having impairments with different subprocesses, e.g., acoustic processing (Kries et al., 2023), phonological, semantic or syntactic processing. That being said, the group analyses conducted for this study may obscure individual links between behavioral and neural outcome measures. Studies with more narrowly recruited IWA that have severe phoneme identification and phonological deficits may help getting a better understanding of the current results.

Our third research question emerged post hoc, because we wanted to explore what might underlie changes in phonetic decoding duration/strength in aphasia. We found that at a few phoneme positions, the TG decoding matrices differed between groups (fig.4). Most markedly, the word-initial phoneme's features – generally the phoneme with the highest uncertainty about lexical identity – were maintained longer/"reactivated" in decoders relatively early after phoneme onset in healthy controls, but not in IWA. This supports the idea that IWA may not keep phonetic features of the word-initial phoneme encoded for long enough, such that integration with higher levels of processing may be hampered. Phonetic decoding for the second phoneme position (relative to word onset and offset) also showed a group difference (fig.S.5), which was similar to the results when taking into account all phonemes, i.e., altered strength of encoding around the matrix diagonal. The first and second phoneme positions (and the last and second to last ones) were tested on around 3,000 phoneme trials, whereas the amount of trials beyond these positions sharply drops off (to below 2,200 trials per position, see table 2). This makes the decoding matrices noisier and the group comparisons more unreliable. Future studies that want to employ an analysis by phoneme position should thus make sure to have enough phoneme trials per position.

In addition to analyzing phonetic decoding for different positions relative to word boundaries, we also assessed how lexical entropy affects phonetic encoding in IWA and controls. We observed that controls, but not IWA, encoded phonemes with high lexical uncertainty for longer period of time than phonemes with low lexical uncertainty. Our findings in healthy controls (older adults) align with Gwilliams et al. (2022)'s results in young adults, suggesting that the ability to flexibly adjust encoding duration based on lexical uncertainty may be crucial. IWA lack this adaptive encoding mechanism, which might thus lead to language processing impairments. IWA appear unable to encode phoneme properties for a sufficient duration to resolve lexical identity. This shows that integration between different levels of processing might be maladaptive in aphasia.

One of the main differences between the Gwilliams et al. (2022) study and the current one is participant age, i.e., younger adults versus older adults. Is it possible that we were not able to reproduce the word-initial phoneme's pre-activation (supplementary fig.S.6), because older adults (with or without aphasia) process prediction-related cues differently than younger adults? Gillis and Kries et al. (2023) showed that a linguistic encoding model, including a measure of surprisal, decreases with advancing age. With regard to word-initial phoneme feature decoding, it is possible that while younger adults pre-activate phonetic features as a results of word prediction (Gwilliams et al., 2022), older adults might instead reactivate features and maintain them longer. However, we can only speculate about this, since the study by Gwilliams et al. (2022) and the present study have also considerable methodological differences other than age, i.e., MEG versus EEG, 2 hours story listening versus 25 minutes, different stimuli, different language, etc. Nonetheless, these are interesting hypotheses to test in a dataset with healthy younger and older adults in the future.

In the study by (Gwilliams et al., 2022), the authors demonstrate that the duration for which phonetic features are maintained in the same neural population (0.08 s) matched with and is contingent on the duration of the phoneme input, i.e., 0.08 s. In our study, we found that the duration of maintaining phonetic information (0.14 s) is similar to the average phoneme duration in the stimulus (0.10 s), thus is in line with previous findings. Recently, Gwilliams et al. (2024) further established the link between stimulus input time and generalization time of the neural representation by applying TG analyses not only at phoneme level, but also at word, phrase and sentence level. They observed that the dynamic coding scheme found at phoneme level may be a more general principle underlying sensory-cognitive processing, and that it is hierarchical in nature, referred to as hierarchical dynamic coding (HDC) (Gwilliams et al., 2024). In the current study, given that IWA encode phonetic features less strongly and more diffusely, it could potentially be a pattern that is observable in IWA at other language processing levels, too. If neurotypical healthy adults with intact speech comprehension show a specific coding pattern, while individuals with a language impairment show a different pattern, then maybe the neurotypical coding pattern is conditional for successful speech comprehension. It is possible that a certain spatio-temporal concentration in neural encoding of speech features (for example phonetic features) is necessary in order to efficiently integrate the information with other language representation levels.

The current paper has some limitations. As already mentioned, we recruited a heterogeneous sample of IWA. While 39 participants with aphasia is a substantial number, especially relative to the general aphasia literature, it remains a small sample and it would be important to see if our findings can be reproduced in a different aphasia dataset, with longer recordings than 25 minutes of story listening, it would also be interesting to look at word- or phrase-level representations. Furthermore, the lesion site and its effect on neural recordings pose a problem for more detailed interpretation of results. Therefore, it would be essential to record data from a control group that has a right-sided lesion without aphasia in the future, so as to investigate to what extent altered encoding of speech is related to structural versus functional changes in aphasia. It is generally an issue in aphasia studies that control groups most of the time consist of healthy adults, whereas a comparison with demographically-matched adults with a brain lesion without aphasia would present a more precise control group to investigate language impairments. It could also be interesting to simulate aphasia in neurotypical adults by distorting speech stimuli, e.g., by using vocoded speech or faster speech. This may also help to get insight into whether altered speech encoding is related to structural changes after stroke or due to functional impairments. Another limitation of the approach we employed here is that it is not reliable at an individual participant level, at least not with the relatively limited amount of data available.

4 Conclusion

Overall, we found that phonetic processing is robustly encoded in EEG responses of older adults. The primary marker of aphasia was altered strength of phonetic encoding after initial processing, rather than a slower evolution of neural representation or the absence of encoding. Moreover, phonemes with high uncertainty about word identity were encoded longer in controls than in individuals with aphasia, indicating that encoding phonetic information until word identity is resolved may be a crucial mechanism for successful speech comprehension. These results suggest that speech impairments in aphasia may be driven by difficulties maintaining lower-order information long enough to recognize lexical items.

5 Methods and methods

5.1 Participants

We tested 39 IWA in the chronic phase after stroke ($\geq$ 6 months) and 24 healthy controls that were age-matched at group-level. For more information on how IWA were recruited, please see Kries et al. (2024). We only included IWA that had no formal diagnosis of a psychiatric or neurodegenerative disorder and that had a left-hemispheric or bilateral lesion. All aphasia participants were tested in the chronic phase after stroke (time since stroke onset in months (median(range)): 16.1(6-126.1). The aphasia sample was checked for language impairments at the moment of data collection using two standardized diagnostic aphasia tests, i.e., the diagnostic test ScreeLing Visch-Brink et al. (2010) and the Dutch picture-naming test (Nederlandse Benoemtest (NBT); Van Ewijk et al. (2020), using the same procedure as reported in Kries et al. (2024); results depicted in fig.1A; individual scores are available in Kries et al. (2024)). On both tests, the aphasia group had significantly lower results than the control group (ScreeLing cut-off threshold: 68/72 points; NBT cut-off threshold: 255/276 points; ScreeLing(mean(SD)): aphasia: 61.5(9.4), control: 69.9(2.5), $p<0.001$; NBT(mean(SD)): aphasia: 224.7(58.1), control: 271.3(4.1), $p<0.001$).

All participants were Dutch native speakers from Flanders, Belgium. Age, sex, education, handedness and multilinguality did not differ between groups (see Kries et al. (2024) for more details). Demographic information was acquired via a self-reported questionnaire. Handedness was assessed via the Edinburgh Handedness Inventory (Oldfield, 1971). Details about the stroke in IWA, i.e., time since stroke onset, stroke type, occluded blood vessel, lesion location and speech-language therapy, can be found in Kries et al. (2024). To visualize the damaged brain tissue of IWA, a lesion overlap image was created (supplementary fig. S.1). The largest lesion overlap was present in left inferior frontal gyrus. Informed consent was obtained from all participants for the recruitment via screening and for the data collection in the chronic phase. The study received ethical approval by the medical ethical committee of KU Leuven and UZ Leuven (S60007) and is in accordance with the declaration of Helsinki.

5.2 Experimental design

The EEG measurements were conducted in a soundproof room with a Faraday cage using a 64-channel EEG system (ActiveTwo, BioSemi, Amsterdam, NL) at 8192 Hz sampling frequency. Participants listened to a 25-minute story, "De wilde zwanen" (The Wild Swans), written by Hans Christian Andersen, narrated by a female Flemish-native speaker, while seated one meter away from a screen, focusing on a fixation cross to minimize eye movement artifacts. The story was divided into 5 parts, approximately 4.8 minutes each, with silences reduced to 200 ms and a sample rate of 48 kHz. Stimuli were presented binaurally via shielded ER-3A insert earphones (Etymotic Research) at 60 dB SPL (A weighted), adjusted if necessary based on hearing thresholds (see Kries et al. (2024) for more details). Following each story part, participants answered yes/no and multiple-choice questions to ensure attentiveness (see Kries et al. (2024) for results of behavioral questions).

5.3 EEG signal processing

The EEG signal processing was conducted using MATLAB (version 9.1.0.441655 (R2016b)). Data from the 5 story parts (epochs) were concatenated, followed by eye movement artifact removal using the multichannel Wiener filter (Somers et al., 2018). The EEG signal was referenced to the common average. For high-pass filtering, we applied a least squares filter with filter order of 2000, passband frequency of 0.5 Hz and stopband frequency of 0.45 Hz. For low-pass filtering, we employed a least squares filter with a filter order of 2000, passband frequency of 25 Hz and stopband frequency of 27.5 Hz. Subsequently, EEG data was downsampled to 128 Hz and normalized by subtracting the mean and dividing by the standard deviation (SD). Next, the EEG data was epoched, centered around phoneme onsets, specifically from -0.2 s before phoneme onset to 0.6 s after onset. The story contained 13560 phonemes and thus also the same amount of epochs. No baseline correction was applied during the epoching process.

5.4 Modelled features

The stimulus narrative was annotated for different properties at the phoneme level, namely 18 phonetic features, phoneme position relative to word onset and relative to word offset, as well as lexical entropy. To get there, first an aligner (Duchateau et al., 2009) was used to create alignment files containing the identity and timing of each phoneme and each word of the stimulus. We annotated the phonemes for 18 phonetic features, based on the multi-value feature system reported in (King and Taylor, 2000), which features voicing, manner of articulation, place of articulation, roundness and front-backness. **Voicing** indicates whether the vocal cords vibrate during speech production. For example, during a /b/ sound, the vocal cords vibrate, but not during a /p/, voicing thus has 2 features, voiced and unvoiced. **Manner** describes how air flows through the articulators during speech production. We tested 6 manner features, namely nasal, fricative, occlusive and approximant for consonsants, and short vowel and long vowel

for vowels. **Place** refers to the positioning of the articulators (e.g., teeth, tongue, lips) during speech production. We tested 8 place features, namely dental, coronal, glottal, labial and velar for consonsants, and low, mid, high for vowels. **Roundness** refers to rounding of the lip, for example during a /w/ sound the lips are rounded, but not during a /f/ sound. Roundness thus has 2 features, rounded and unrounded. **Front-backness** refers to the position of the tongue in the mouth, entailing the features front, central and back. However before further analysis, we removed phonetic features that occurred less than 5% of total phonemes in the stimulus, leading to removal of 3 place features, namely dental, glottal and low vowel. This led to 18 features that were used for further analysis.

Subset variables. To investigate whether phonemes are differentially encoded depending on their phoneme position within words and their lexical entropy, we subset phonemes based on these properties. **Phoneme position** relative to word onset was coded as 1 for the first phoneme within words, 2 for the second phoneme within words, etc. until the 5th position, and the opposite order for phoneme position relative to word offset. This means that there were less phoneme occurrences per position, as evidenced in table 2. Concretely, this was executed by using the word onsets to chunk the phoneme onsets by word, then loop through the chunks and serialize the phoneme onsets within word. **Lexical entropy** measures the level of competition among words that match the current phoneme input. For example, upon hearing the sounds /pl/, many potential words are activated (n=999), indicating a high degree of competition. As additional phonemes are heard, the number of possible words diminishes (e.g., for /plu/, n=162, and for /plur/, n=19), resulting in a lower degree of competition. This competition level is calculated using the Shannon entropy of the words in the activated cohort (Gwilliams and Davis, 2022). The first phoneme of each word was excluded from this analysis. Entropy was calculated based on the SUBTLEX-NL database (Keuleers et al., 2010) and a custom pronunciation dictionary. Then we computed the 33rd and 66th percentiles to split entropy values into 3 equal parts and used the third with the lowest and the one with the highest values for the decoding analysis. In table 2, you can see how many phoneme occurrences were present in the thirds.

Table 2: **Phoneme occurrences for subset variables**

level	**n**	**meaning**
phoneme position relative to word onset/offset		
1	3002	first/last phoneme within word
2	2948	second (to last) phoneme within word
3	2199	third (to last) phoneme within word
4	1524	fourth (to last) phoneme within word
5	1187	fifth (to last) phoneme within word
lexical entropy		
1	3427	phonemes with low entropy
3	3592	phonemes with high entropy

5.5 Decoding analysis

We decoded each phonetic feature from the 64-sensor EEG signal. Specifically, we decoded phonetic features using a TG approach. TG entails testing whether a temporal decoder trained on data at time t can accurately decode data at time t from a testing set. Instead of assessing decoding accuracy solely at the specific time point the model was trained on, we evaluated its accuracy across all possible train/test time combinations. TG is represented by a square decoding matrix of training time vs. testing time. In practice, we fit logistic regressions using 5-fold cross-validation. We assessed the decoding performance by computing the area under the curve (AUC) of the receiver-operating curve (ROC).

To assess for how long phonetic features can be decoded from brain data, we extracted the temporal decoders where train and test time were identical, i.e., the diagonal of the TG matrix. Then we tested at what time points the decoding performance is above chance. We also tested whether there is a cluster(s) of the TG matrix that is decodable above-chance. This allows us to investigate for how long information captured at a any given temporal decoder is represented or generalized, namely by calculating the width of the above-chance cluster of the matrix. In other words, we calculated the duration of above-chance generalization across temporal decoders. Additionally, to explore the dynamics of neural representations, we compared the mean duration of above-chance generalization across temporal decoders to the duration of above-chance temporal decoding (i.e., comparing the diagonal of the matrix to its rows). These metrics were examined within each participant and analyzed using second-level statistics across participants.

To investigate which EEG sensors are most important for phonetic decoding, we decoded features from each sensor individually, thus passing the epochs' time course of a given sensor to the decoder. This can tell us how much each sensor contributes across time to phonetic feature decoding.

5.6 Comparing performance between trial subsets

To assess decoding performance across theoretically interesting subsets of trials (such as start vs. end of a word or high vs. low entropy), we slightly changed our train/test cross-validation procedure. The classifier was trained using the full training dataset (80% of 13560 phonemes), while the test set was divided according to the different levels of interest (table 2). We then independently evaluated the model's performance on each subset of the test data, yielding a time-course or generalization matrix for each group of trials being analyzed.

5.7 Statistical analysis

To compare TG matrices and its diagonals against chance-level, we used permutation-based statistics. This was performed on the group-level average decoding performance using a one-sample permutation cluster test implemented using the MNE Python package (Gramfort et al., 2013). First, a t-value is

calculated for every data point of the TG matrix. To determine the statistical significance, the data undergoes 10,000 iterations of random permutation, wherein the signs of the data points are randomly flipped. For each permutation, clusters of neighboring data points are identified, and the sum of the t-values within these clusters is computed. This generates a null distribution of summed t-values from the permuted data. The significance of the observed effect is then assessed by comparing the observed cluster's t-value sum to this null distribution. If the observed sum is greater than 95% of the sums from the null distribution, the effect is considered statistically significant, indicating that it is unlikely to have occurred by chance.

To compare the aphasia and control group, we used the permutation cluster test as implemented in the MNE Python package (Gramfort et al., 2013). This test compares groups by first calculating differences at each data point of the TG matrices. It then shuffles data between groups 10'000 times to create a null distribution. Significant clusters, where differences are consistent, are identified by comparing observed values to those in the null distribution. If observed values rank in the top 5% ($p<.05$) of the null distribution, the test indicates statistically significant differences between groups.

To test whether the TG matrix pattern is dynamically evolving (i.e., the diagonal is larger than the width of the decoding pattern), we (1) extracted the diagonal and horizontal (average across y-axis) time courses per participant, (2) subtracted the diagonal from the horizontal time course for each participant, and (3) used a one-sample permutation cluster test to test whether this delta is significantly different from zero.

To analyze topographical differences of phonetic feature decoding between the aphasia group and the control group, we submitted the decoding performance of each sensor across subjects and groups to a mass-univariate independent samples t-test (Brodbeck, 2020).

Generative AI

During the preparation of this work, the first author used ChatGPT in order to enhance readability. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Ahlfors, S. P., Han, J., Belliveau, J. W., and Hämäläinen, M. S. (2010). Sensitivity of meg and eeg to source orientation. *Brain topography*, 23:227–232.

Blumstein, S. E. (1998). Phonological aspects of aphasia. *Acquired aphasia*, 3:157–186.

Brodbeck, C. (2020). Eelbrain 0.32. `http://doi.org/10.5281/zenodo.3923991`.

Cassidy, J. M., Wodeyar, A., Wu, J., Kaur, K., Masuda, A. K., Srinivasan, R., and Cramer, S. C. (2020). Low-frequency oscillations are a biomarker of injury and recovery after stroke. *Stroke*, pages 1442–1450.

Cohen, R., Abboud, S., and Arad, M. (2015). Monitoring brain damage using bioimpedance technique in a 3d numerical model of the head. *Medical Engineering & Physics*, 37(5):453–459.

De Clercq, P., Kries, J., Vanthornhout, J., Gerrits, R., Francart, T., and Vandermosten, M. (2024). Neural substrates and behavioral relevance of speech envelope tracking: evidence from post-stroke aphasia. *bioRxiv*.

Duchateau, J., Kong, Y. O., Cleuren, L., Latacz, L., Roelens, J., Samir, A., Demuynck, K., Ghesquière, P., Verhelst, W., and Van hamme, H. (2009). Developing a reading tutor: Design and evaluation of dedicated speech recognition and synthesis modules. *Speech Communication*, 51(10):985–994.

Gilbert, C. D. and Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron*, 54(5):677–696.

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

Goldenholz, D. M., Ahlfors, S. P., Hämäläinen, M. S., Sharon, D., Ishitobi, M., Vaina, L. M., and Stufflebeam, S. M. (2009). Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography. *Human Brain Mapping*, 30(4):1077–1086.

Gramfort, A., Luessi, M., Larson, E., Strohmeier, D. A. E. D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., and Hämäläinen, M. S. (2013). Meg and eeg data analysis with mne-python. *Frontiers in Neuroscience*, 7:1–13.

Gwilliams, L. and Davis, M. H. (2022). Extracting language content from speech sounds: the information theoretic approach. In Holt, L. L., Peelle, J. E., Coffin, A. B., Popper, A. N., and Fay, R. R., editors, *Speech Perception*, chapter 5, pages 113–139. Springer Cham, Switzerland.

Gwilliams, L., King, J. R., Marantz, A., and Poeppel, D. (2022). Neural dynamics of phoneme sequences reveal position-invariant code for content and order. *Nature Communications*, 13.

Gwilliams, L., Linzen, T., Poeppel, D., and Marantz, A. (2018). In spoken word recognition, the future predicts the past. *Journal of Neuroscience*, 38(35):7585–7599.

Gwilliams, L., Marantz, A., Poeppel, D., and King, J.-R. (2024). Hierarchical dynamic coding coordinates speech comprehension in the brain. *bioRxiv*.

Harnish, S. M. (2018). Anomia and anomic aphasia: Implications for lexical processing. *The Oxford handbook of aphasia and language disorders*, pages 121–144.

Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(May):393–402.

Keuleers, E., Brysbaert, M., and New, B. (2010). Subtlex-nl: A new measure for dutch word frequency based on film subtitles. *Behavior research methods*, 42(3):643–650.

King, J.-R., Gwilliams, L., Holdgraf, C., Sassenhagen, J., Barachant, A., Engemann, D., Larson, E., and Gramfort, A. (2020). Encoding and Decoding Framework to Uncover the Algorithms of Cognition. In *The Cognitive Neurosciences*. The MIT Press.

King, S. and Taylor, P. (2000). Detection of phonological features in continuous speech using neural networks. *Computer Speech and Language*, 14(4):333–353.

Koen, J. D. and Rugg, M. D. (2019). Neural dedifferentiation in the aging brain. *Trends in Cognitive Sciences*, 23(7):547–559.

Kries, J., Clercq, P. D., Lemmens, R., Francart, T., and Vandermosten, M. (2023). Acoustic and phonemic processing are impaired in individuals with aphasia. *Scientific Reports*, 13(1):11208.

Kries, J., De Clercq, P., Gillis, M., Vanthornhout, J., Lemmens, R., Francart, T., and Vandermosten, M. (2024). Exploring neural tracking of acoustic and linguistic speech representations in individuals with post-stroke aphasia. *Human Brain Mapping*, 45(8):e26676.

Levy, D. F. and Wilson, S. M. (2019). Categorical encoding of vowels in primary auditory cortex. *Cerebral Cortex*, 30:618–627.

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, 343.

Mesgarani, N., David, S., and Shamma, S. (2007). Representation of phonemes in primary auditory cortex: How the brain analyzes speech. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE.

Oldfield, R. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1):97–113.

Park, W., Kwon, G. H., Kim, Y. H., Lee, J. H., and Kim, L. (2016). Eeg response varies with lesion location in patients with chronic stroke. *Journal of NeuroEngineering and Rehabilitation*, 13.

Piastra, M. C., Oostenveld, R., Schoffelen, J. M., and Piai, V. (2022). Estimating the influence of stroke lesions on meg source reconstruction. *NeuroImage*, 260.

Purcell, J. J., Wiley, R. W., and Rapp, B. (2019). Re-learning to be different: Increased neural differentiation supports post-stroke language recovery. *NeuroImage*, 202:116145.

Salinet, A. S. M., Panerai, R. B., Robinson, T. G., and Salinet, A. (2014). The longitudinal evolution of cerebral blood flow regulation after acute ischaemic stroke e x t r a. *Cerebrovasc Dis Extra*, 4:186–197.

Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., and Fedorenko, E. (2020). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, 138:1–52.

Sigman, M. and Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind's bottleneck. *PLOS Biology*, 3(2):null.

Somers, B., Francart, T., and Bertrand, A. (2018). A generic eeg artifact removal algorithm based on the multi-channel wiener filter. *Journal of Neural Engineering*, 15.

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, Houten, NL.

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

Vorwerk, J., Cho, J. H., Rampp, S., Hamer, H., Knösche, T. R., and Wolters, C. H. (2014). A guideline for head volume conductor modeling in eeg and meg. *NeuroImage*, 100:590–607.

Wilson, S. M., Bautista, A., and McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *NeuroImage*, 171:62–74.

Wilson, S. M., Entrup, J. L., Schneck, S. M., Onuscheck, C. F., Levy, D. F., Rahman, M., Willey, E., Casilio, M., Yen, M., Brito, A. C., Kam, W., Davis, L. T., Riesthal, M. D., and Kirshner, H. S. (2023). Recovery from aphasia in the first year after stroke. *Brain*, 146:146–1021.

Yi, H. G., Leonard, M. K., and Chang, E. F. (2019). The encoding of speech sounds in the superior temporal gyrus. *Neuron*, 102:1096–1110.

Zbesko, J. C., Nguyen, T.-V. V., Yang, T., Frye, J. B., Hussain, O., Hayes, M., Chung, A., Day, W. A., Stepanovic, K., Krumberger, M., Mona, J., Longo, F. M., Doyle, K. P., and manuscript, N. D. A. (2018). Glial scars are permeable to the neurotoxic environment of chronic stroke infarcts hhs public access author manuscript. *Neurobiol Dis*, 112:63–78.

Zhang, H. and Hinzen, W. (2022). Grammar in 'agrammatical' aphasia: What's intact? *PLoS ONE*, 17.

Supplementary material

S.1 Lesion information

MRI data collection was not a part of the planned data collection for this study, but ethical agreement was given to access the scans from the medical files of the university hospital. Thus, we had access to T2-weighted FLAIR images that were administered in the acute stage post-stroke, based on which the segmentations of the affected stroke tissue were manually performed. For 10 participants, scans were available from the chronic phase post-stroke, as they took part in an MRI study (De Clercq et al., 2024). 12 participants had participated in an earlier aphasia study, and thus we could use the already established segmentation maps.

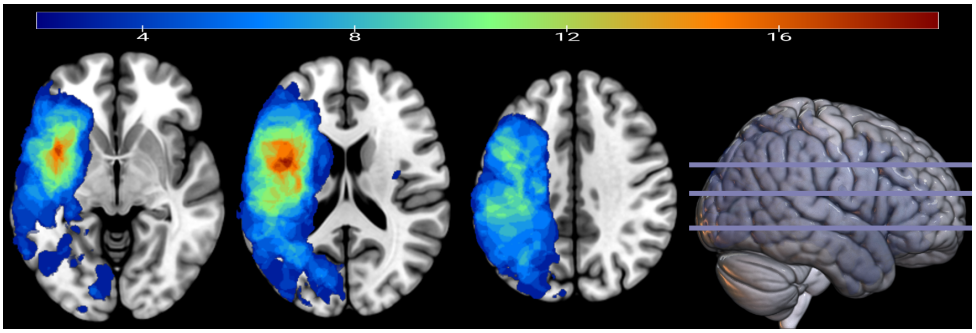

Figure S.1: **Lesion overlap image of the aphasia sample.** The maximum overlap corresponds to 19 out of the total sample of 39 individuals with aphasia. Axial slices are shown in neurological orientation.

S.2 Group demographics and behavioral variables

In supplementary table S.1, we summarized demographic information by group (details can be found in supplementary table S.2). Age, sex, education, handedness and multilinguality did not differ between groups (age: W = 464, p = 0.96; sex: $\chi^2 = 0$, df = 1, p = 1; education: $\chi^2 = 7.26$, df = 4, p = 0.1; handedness: $\chi^2 = 0.063$, df = 2, p = 0.98; multilingual: $\chi^2 = 0.182$, df = 1, p = 0.66). Details about the stroke in IWA, i.e., time since stroke onset, stroke type, occluded blood vessel, lesion location and speech-language therapy, can be found in supplementary table S.2. To visualize the damaged brain tissue of IWA, a lesion overlap image was created (supplementary fig. S.1). Demographic information was acquired via a self-reported questionnaire. Handedness was assessed via the Edinburgh Handedness Inventory (Oldfield, 1971).

Table S.1: Demographics, language-diagnostic information and covariates by group.

		Control	**Aphasia**	
Demographics				
Age in years (mean(SD))$^{\diamond}$		71.5(7)	69.5(12.4)	n.s.
Sex (n(%))	female	8(33.3%)	13(33.3%)	n.s.
	male	16(66.7%)	26(66.7%)	
Education (n(%))	primary education	0(0%)	3(7.7%)	n.s.
	secondary education	5(20.8%)	15(38.4%)	
	tertiary education/college	8(33.3%)	10(25.6%)	
	Master's degree	9(37.5%)	11(28.2%)	
	PhD	2(8.3%)	0(0%)	
Handedness (n(%))	right	22(91.6%)	35(89.7%)	n.s.
	left	1(4.2%)	2(5.1%)	
	ambidextrous	1(4.2%)	2(5.1%)	
Multilingual (n(%))	yes	22(91.7%)	33(84.6%)	n.s.
	no	2(8.3%)	6(15.3%)	
Language diagnostic tests				
ScreeLing (max=72) (mean(SD))		69.9(2.5)	61.5(9.4)	* * *
Picture naming (max=276) (mean(SD))		271.3(4.1)	224.7(58.1)	* * *
Covariates				
Hearing (Fletcher index in dB HL) (mean(SD))		24.5(12)	28.4(14.1)	n.s.
Cognition (OCS composite score in %) (mean(SD))		94.2(4.7)	80.4(16.6)	* * *
Alertness (Likert scale) (median(range))		4.75(3.5-5)	3.5(2-5)	**
Fatigue (Likert scale) (median(range))		1.5(1-4.5)	3.5(1-4.5)	**

$^{\diamond}$ Age is also used as a covariate; n.s. = not significantly different; p<0.05*; p<0.01**; p<0.001***

S.3 Demographic and lesion information of individuals with aphasia

Table S.2: Demographic, lesion and diagnostic information of the aphasia group

ID(n=39)	age	sex	Time since stroke (months)	Stroke type	Blood vessel	Lesioned hemi-sphere	SLT	NBT score (max=276 cutoff=255)	ScreeLing score (max=72 cutoff=68)
sub-006	86	m	22.8	ischemia	VA	bilateral	yes	263	70
sub-008	71	f	21.1	ischemia	MCA	left	no	273	69
sub-009	67	m	22.6	ischemia	MCA	left	no	262	65.5
sub-010	74	m	20.6	ischemia	PCA	left	no	◇	69
sub-014	75	m	18.2	ischemia	MCA	left	yes	185	58
sub-016	68	m	18.0	ischemia	MCA	bilateral	yes	265	69
sub-017	88	m	8.2	ischemia	MCA/ PCA	bilateral	no	245	56
sub-018	61	f	29.4	ischemia	MCA	left	yes	263	72
sub-019	72	m	8.6	ischemia	PCA	left	◇	211	51
sub-020	42	m	18.7	ischemia	MCA	left	yes	273	71
sub-021	81	m	8.6	hemorrhage	◇	left	no	252	70
sub-022	78	f	8.2	hemorrhage	◇	left	yes	207	56
sub-024	69	m	6	ischemia	MCA	left	yes	260	68.5
sub-025	69	m	11.5	ischemia	MCA	left	yes	197	58.5
sub-026	71	m	31	ischemia	MCA	left	yes	250	69
sub-027	76	m	126.2	ischemia	MCA	left	yes	254	63
sub-028	80	m	25.2	ischemia	◇	left	no	195	54
sub-029	75	m	22.6	ischemia	MCA	left	yes	193	55
sub-030	49	m	13.1	hemorrhage	◇	left	yes	217	57
sub-031	79	m	12.9	ischemia	MCA	left	yes	263	63
sub-032	76	m	94.8	ischemia	MCA	left	yes	3	28
sub-034	79	m	13.4	ischemia	◇	bilateral	yes	244	69
sub-035	64	f	31.5	ischemia	MCA/ ACA	left	yes	276	71
sub-049	85	f	8.3	ischemia	PICA	left	yes	242	65.5
sub-050	81	f	8.7	ischemia	MCA	left	yes	134	53
sub-052	72	m	120.7	ischemia	MCA	left	yes	175	48
sub-053	41	f	6.1	ischemia	MCA	left	yes	271	70
sub-054	69	m	17.2	ischemia	MCA	left	yes	268	69
sub-056	76	m	14.9	ischemia	MCA	left	yes	218	58
sub-059	86	m	15.2	ischemia	ICA	left	no	196	51
sub-060	70	m	15.4	ischemia	MCA	left	yes	150	◇
sub-061	54	f	12.7	ischemia	MCA	left	yes	276	72
sub-063	63	f	13.9	ischemia	MCA	left	yes	272	69.5
sub-064	81	f	12.2	ischemia	crypto.	bilateral	yes	251	62.5
sub-065	50	f	23.8	ischemia	MCA	left	yes	93	44
sub-067	40	f	38.5	ischemia	MCA	left	yes	164	56
sub-068	59	m	12.2	ischemia	MCA	left	no	247	63.5
sub-069	65	m	17.1	ischemia	MCA	left	no	258	65
sub-070	71	f	16.1	ischemia	MCA	bilateral	no	273	59.5
Total	-	26 m 13 f	-	34 ischemia 3 hemorrhage	27 MCA 3 PCA 1 ACA 1 VA 1 PICA 1 ICA	32 left 5 bilateral	28 yes 10 no		

SLT=speech-language therapy; VA=vertebral artery; MCA=middle cerebral artery; PCA=posterior cerebral artery; ACA=anterior cerebral artery; LLA=lateral lenticulostriate arteries; PICA=posterior inferior cerebellar artery; ICA=internal carotid artery; crypto=cryptogenic; ◇=data not available

S.4 Phonetic feature decoding duration in aphasia

We assessed significance of the decoding time course of each of the 18 phonetic features using a one-sample cluster permutation test across subjects. We found that all 18 phonetic features were decodable above chance in the aphasia group (fig.S.2A; table S.3).

When averaging across phonetic features, we observed above chance decoding from -0.08 to 0.56s ($p<.001$) relative to phoneme onset for the aphasia group.

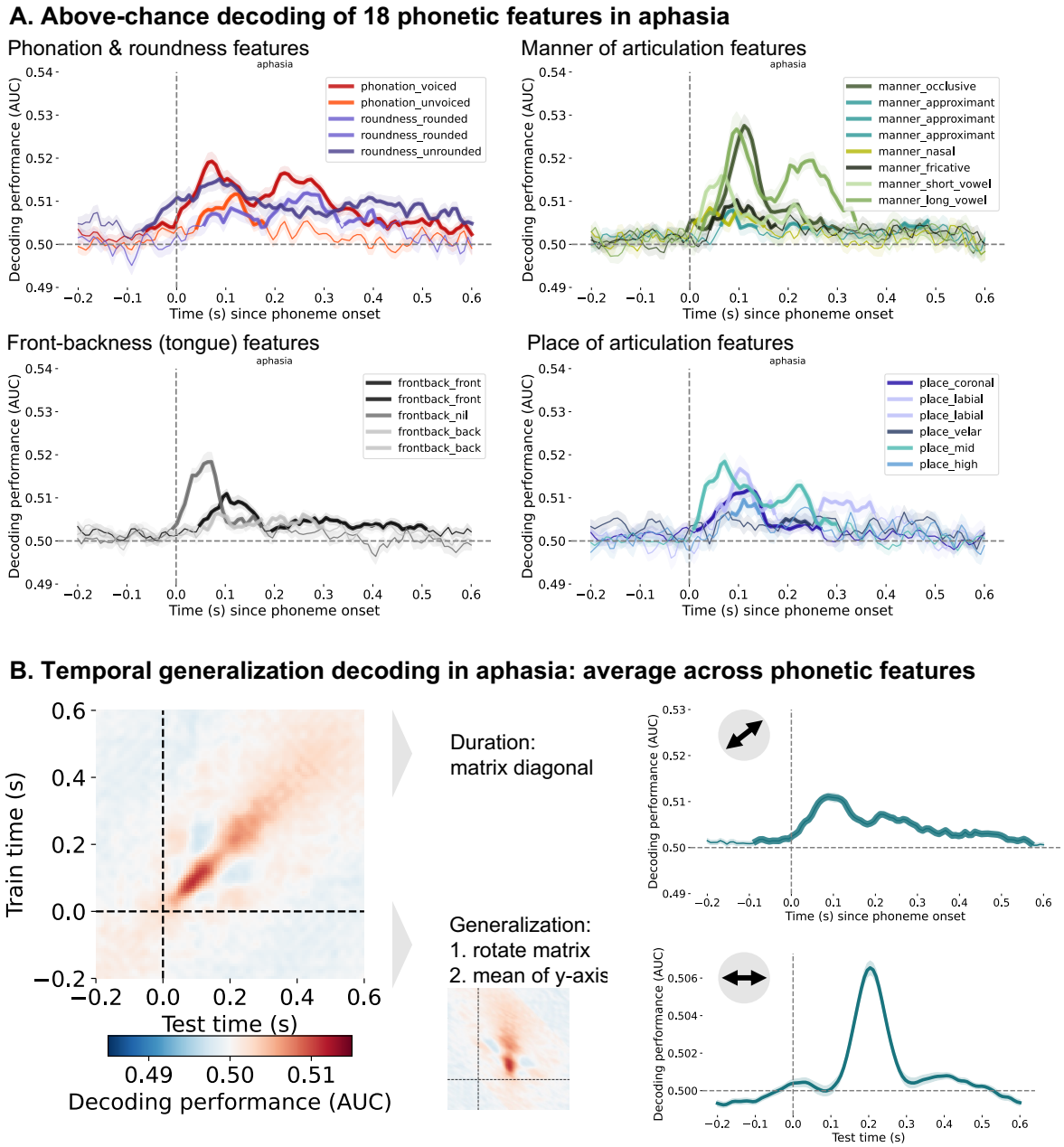

Figure S.2: **A.** Above-chance decoding of 18 phonetic features in aphasia. The features vary in decoding performance and over time. The bold parts of the lines indicate the time points where the features are significantly decoded above chance. **B.** The TG approach in aphasia revealed that phonetic features are decodable above chance for 0.6s (top right panel) and generalizable for 0.13s (bottom right panel). The decoding duration corresponds to the diagonal of the TG matrix, thus the decoding performance when the decoder was trained on the same time slice as was used for testing the decoder. To quantify the generalization time, the TG matrix is rotated and then averaged across the y-axis.

phonetic feature	significant time window (s)	average decoding performance (AUC)	average t-value	p-value
voiced	-0.06 to 0.6	0.508	1.415	<0.001
unvoiced	0.04 to 0.18	0.507	-1.120	<0.001
nasal	0.009 to 0.149	0.506	1.088	<0.001
fricative	0.040 to 0.172	0.506	0.391	<0.001
occlusive	-0.005 to 0.304	0.509	-0.119	<0.001
approximant	0.064 to 0.133	0.505	0.273	0.003
	0.196 to 0.250	0.504	1.459	0.009
	0.429 to 0.483	0.504	1.180	0.009
short vowel	0.017 to 0.250	0.508	-0.509	<0.001
long vowel	0.025 to 0.335	0.514	1.125	<0.001
coronal	0.017 to 0.266	0.506	0.831	<0.001
labial	0.056 to 0.172	0.510	0.972	<0.001
	0.266 to 0.374	0.508	0.598	<0.001
velar	0.188 to 0.242	0.504	3.079	0.005
mid	0.009 to 0.297	0.510	-0.419	<0.001
high	0.087 to 0.141	0.507	-0.957	<0.001
rounded	0.056 to 0.328	0.507	0.886	<0.001
	0.343 to 0.421	0.505	1.078	0.002
unrounded	-0.2 to -0.114	0.505	3.147	0.008
	-0.067 to 0.6	0.509	2.474	<0.001
front	0.048 to 0.172	0.507	2.376	<0.001
	0.196 to 0.506	0.503	1.847	<0.001
central	-0.005 to 0.172	0.508	0.405	<0.001
back	0.056 to 0.172	0.504	2.385	<0.001
	0.188 to 0.297	0.504	2.633	<0.001

Table S.3: Above-chance decoding of phonetic features in individuals with aphasia.

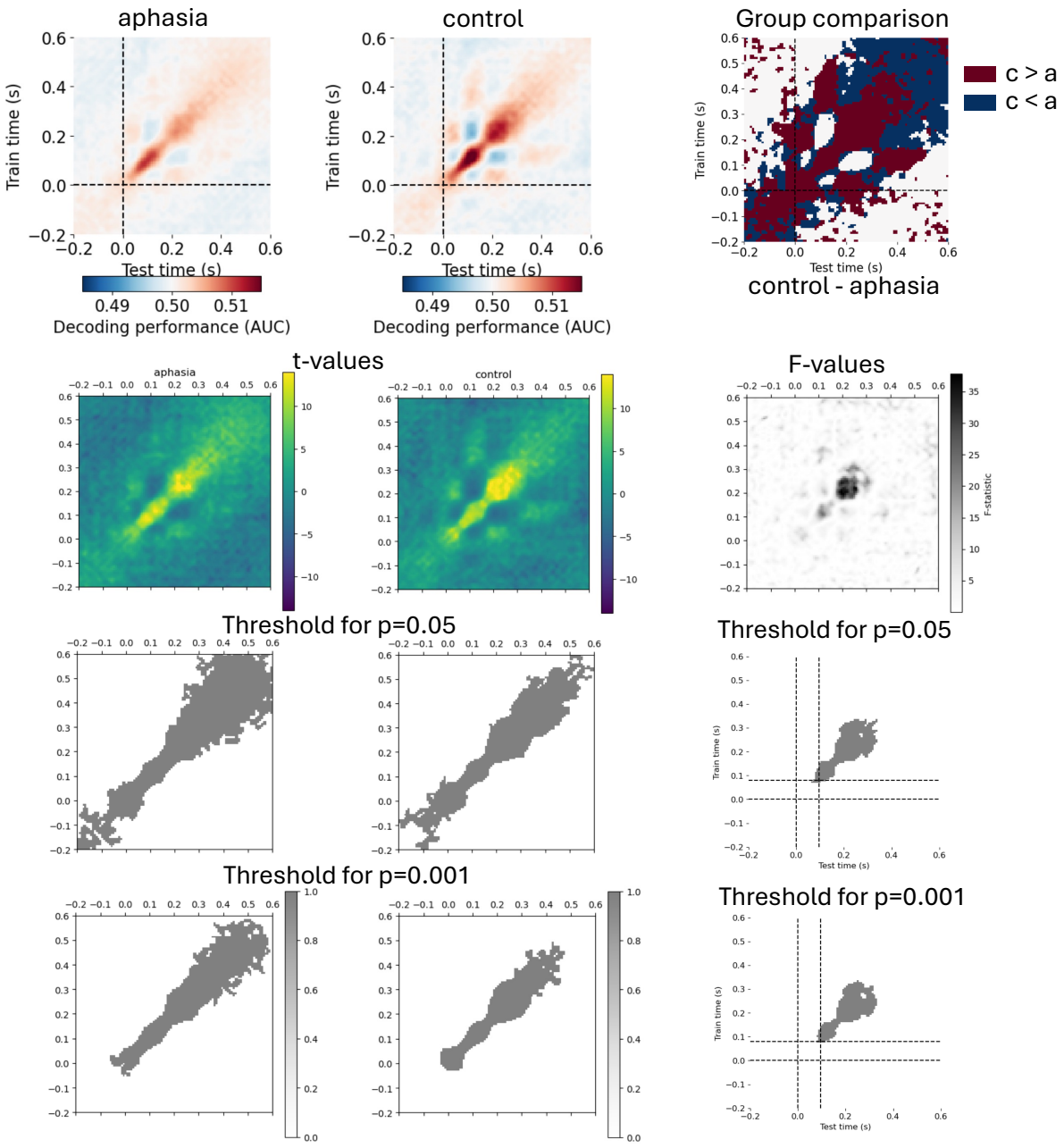

Figure S.3: **Average across phonetic features decoding in aging and aphasia. First row:** decoding performance per group, right side panel: aphasia (a) group subtracted from control (c) group. **Second row:** (one-sample) permutation cluster tests' t-values/F-values. **Third row:** significant clusters for each group and between groups for a cluster-forming p-value threshold of 0.05. **Fourth row:** significant clusters for each group and between groups for a cluster-forming p-value threshold of 0.001

S.5 Duration of above-chance phonetic feature decoding in aphasia and controls

In addition to directly comparing control and aphasia groups, we investigated for how long phonetic features are decodable above chance-level in both groups separately. We used a one-sample permutation cluster test for each group separately.

As indicated in section 2.1, in **controls** we observed above chance decoding from -0.04 to 0.49 s ($p<.001$) relative to phoneme onset (fig.2B top right plot). Thus, on average, phonetic features are decodable for a span of 0.53 s in the control group. In the **aphasia** group, we observed significant above-chance decoding from -0.083 to 0.568 s, thus a span of 0.65 s (fig.3B bold blue line).

In a second step, we compared the above-chance phonetic feature decodability duration between aphasia and control groups. To do so, we 1) extracted the TG matrix diagonal for each individual participant, 2) calculated the standard deviation across the individual time course to compute a confidence interval for each value in the time course, 3) summed up how many time course values' confidence intervals did not contain zero, giving us the amount of significant time lags per participant and 4) conducted an independent samples t-test with 10k permutations to compare the number of significant time lags between the aphasia and control group. This way, we found that the duration of decodability in the aphasia group is not significantly longer than in the control group ($t=-0.066$; $p=0.9$).

S.6 Is the phonetic decoding performance in different peaks associated with behavioral measures within the aphasia group?

We hypothesized that the decoding performance in window 1 would be related to different language processes than the decoding performance in window 2. However, the language-diagnostic, cognitive and anatomical variables that we were primarily interested in did not significantly correlate with the decoding performance in either window. Only age significantly correlated with decoding performance in window 1, specifically the older individuals with aphasia are, the higher the decoding performance was from 0.05s to 0.15s after phoneme onset ($r=0.32$, $p<0.05$).

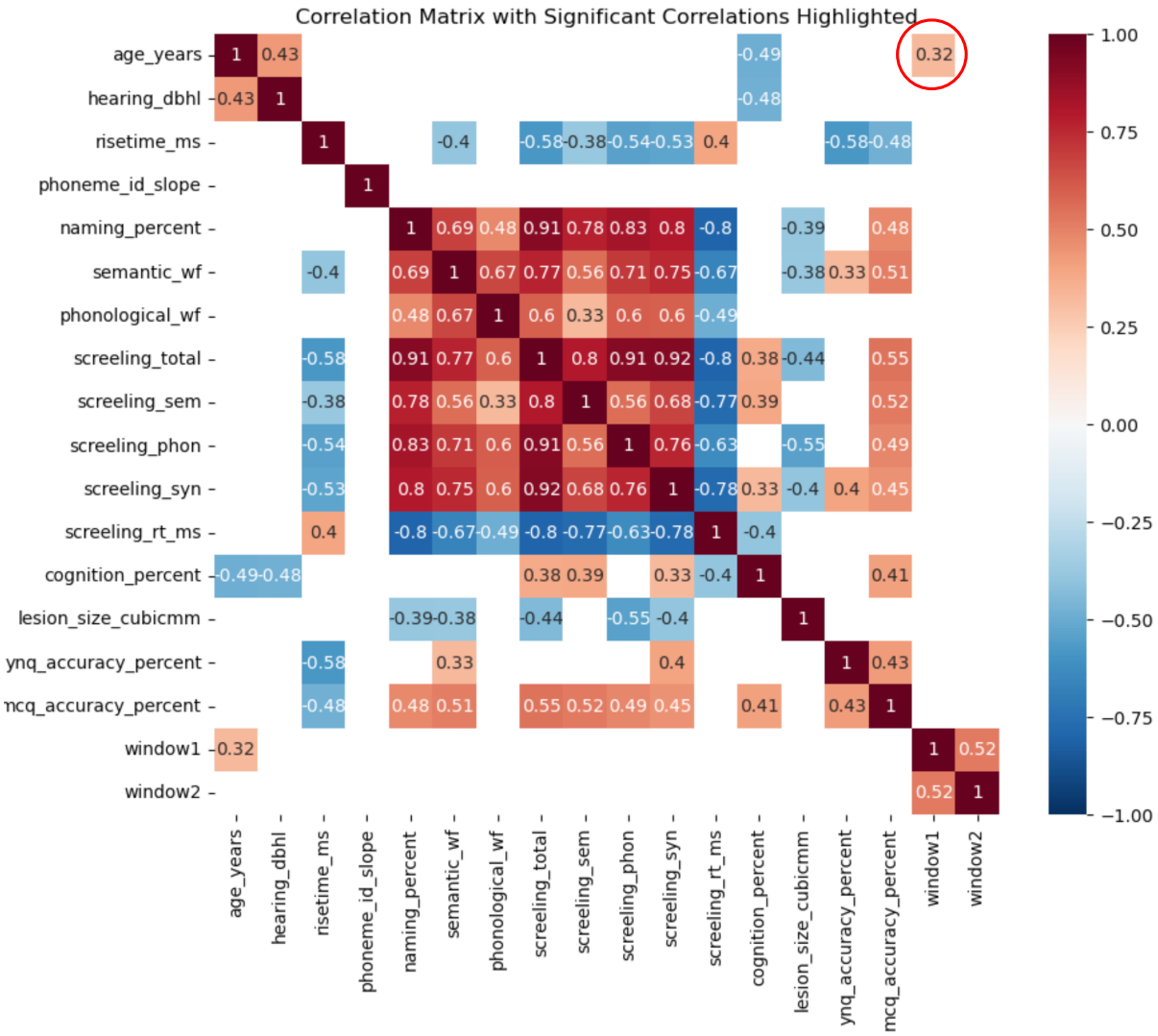

Figure S.4: **Within-aphasia group correlation plot of demographic, language-diagnostic, cognitive, anatomical variables and decoding performance.** We averaged the decoding performance of 2 time windows of interest from the decoding performance time course. Window 1 corresponds to 0.05-0.15s, window 2 corresponds to 0.15-0.3s. The colorbar indicates the correlation value. Only significant correlations are colored.

S.7 Phoneme position analysis

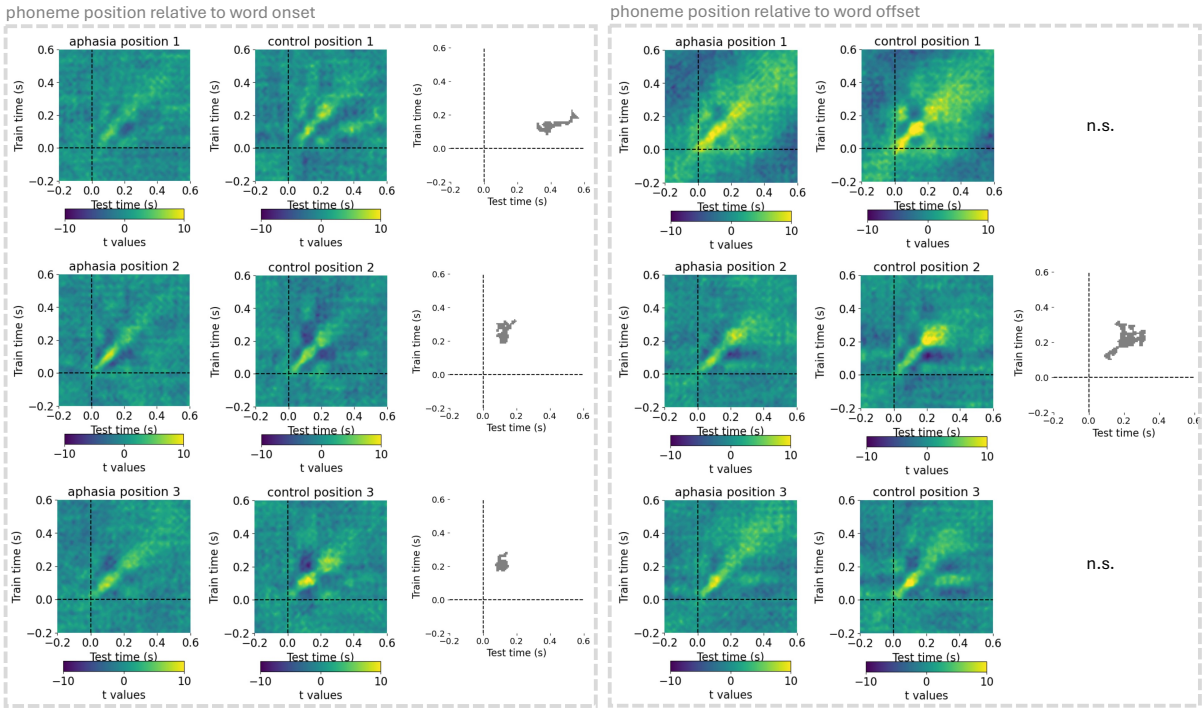

Figure S.5: **Phonetic feature decoding for different phoneme positions relative to word boundaries. Left panel:** TG decoding matrix for first, second and third phoneme positions relative to word onset. **Right panel:** TG decoding matrix for first, second and third phoneme positions relative to word offset. For both panels, the first column shows decoding for the aphasia group, the second column for the control group and third column shows the cluster that was significantly different between groups. n.s.=not significant

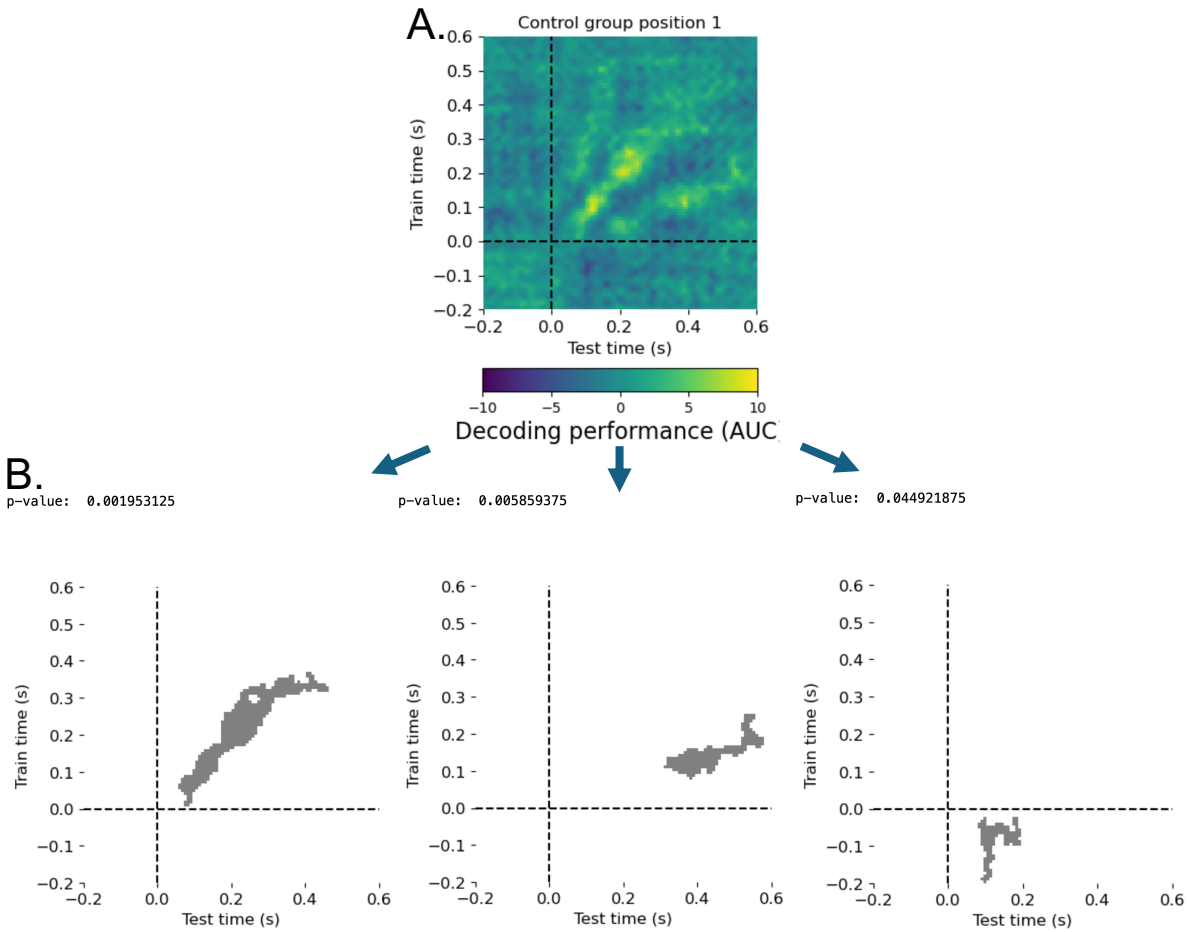

Figure S.6: **Phonetic feature decoding of word-initial phonemes relative to word onset in control group.** To see whether healthy older adults have a similar pre-activation of phonetic features as found in younger adults in a previous study (Gwilliams et al., 2022), we tested the word-initial phoneme position's feature decoding of the control group (**panel A**) against chance. We observed 3 significant clusters (**panel B**), but none of them appears to correspond to a pre-activation, thus above the diagonal. We were not able to replicate a pre-activation of phonetic features of the word-initial phoneme in healthy older adults.

S.8 Entropy analysis

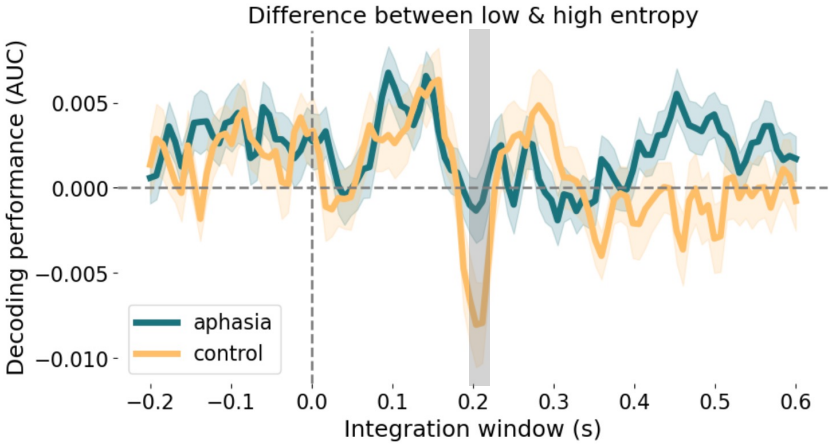

Figure S.7: **Difference between phonetic feature decoding performance of low and high entropy.** The grey bar indicates time lags where the groups significantly differ.