

1 Exploring neural tracking of acoustic and linguistic speech 2 representations in individuals with post-stroke aphasia

3 **Short running title:** Neural speech tracking in post-stroke aphasia

4 Jill Kries^{1,2}, Pieter De Clercq¹, Marlies Gillis¹, Jonas Vanthornhout¹, Robin Lemmens^{3,4,5}, Tom
5 Francart¹ and Maaike Vandermosten¹

6 **Affiliations:**

7 ¹ Experimental Oto-Rhino-Laryngology, Department of Neurosciences, Leuven Brain Institute, KU
8 Leuven, Leuven, Belgium

9 ² Department of Psychology, Stanford University, Stanford, CA, USA

10 ³ Experimental Neurology, Department of Neurosciences, KU Leuven, Leuven, Belgium

11 ⁴ Laboratory of Neurobiology, VIB-KU Leuven Center for Brain and Disease Research, Leuven, Belgium

12 ⁵ University Hospitals Leuven, Department of Neurology, Leuven, Belgium

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

14 **Email addresses:** jill.kries@kuleuven.be / jill.kries@stanford.edu; pieter.declercq@kuleuven.be;
15 marlies.gillis@kuleuven.be; jonas.vanthornhout@kuleuven.be; robin.lemmens@uzleuven.be;
16 tom.francart@kuleuven.be; maaike.vandermosten@kuleuven.be

17 **Acknowledgements:** The authors would like to thank all participants, especially all the brave
18 participants with aphasia and their partners, family or friends that support them. Furthermore, the
19 authors would like to thank Dr. Klara Schevenels for helping with recruitment of participants with
20 aphasia. Thanks also to Dr. Toivo Glatz for methods advice. Moreover, the authors would like to thank
21 everyone who helped with data collection and recruitment: Janne Segers, Rosanne Partoens, Charlotte
22 Rommel, Dr. Ramtin Mehraram, Ines Robberechts, Laura Van Den Bergh, Anke Heremans, Frauke De
23 Vis, Mouna Vanlommel, Naomi Pollet, Kaat Schroeven, Pia Reynaert and Merel Dillen.

24 **Data availability statement:** All data generated or analyzed for this study are available on Open
25 Science Framework (OSF) under the following link: <https://osf.io/6t2ja/>.

²⁶ **Funding Sources:** Research of Jill Kries was supported by the Luxembourg National Research Fund
²⁷ (FNR) (AFR-PhD project reference 13513810). Pieter De Clercq was financially supported by the
²⁸ Research Foundation Flanders (FWO; PhD grant: SB 1S40122N). Research of Marlies Gillis was funded
²⁹ by the FWO (PhD grant: SB 1SA0620N). Research of Jonas Vanthornhout was supported by the FWO
³⁰ (postdoctoral grant: 1290821N). Robin Lemmens is a senior clinical investigator supported by the
³¹ FWO. The presented study also received funding from the European Research Council (ERC) under the
³² European Union's Horizon 2020 research and innovation programme (Tom Francart; grant agreement
³³ No. 637424). Furthermore, this study was financially supported by the FWO grant No. G0D8520N.

³⁴ **Conflict of interest statement:** No conflicts of interest, financial or otherwise, are declared by the
³⁵ authors.

³⁶ **Ethics approval and patient consent statement:** The study received ethical approval by the
³⁷ medical ethical committee of KU Leuven and UZ Leuven (S60007) and is in accordance with the
³⁸ declaration of Helsinki. Informed consent was obtained from all participants for the recruitment via
³⁹ screening and for the data collection in the chronic phase.

⁴⁰ **Author contribution statement:** Conceptualization: MV, JK, TF; Investigation: JK, PDC, MV;
⁴¹ Project Administration: JK; Data Curation: JK; Resources: MV, TF, RL; Methodology: JK, MV, JV,
⁴² MG, PDC, TF; Formal analysis: JK, MG, JV, PDC; Visualization: JK, PDC, MG; Writing – Original
⁴³ Draft Preparation: JK; Writing – Review and Editing: JK, PDC, MG, JV, RL, TF, MV; Supervision:
⁴⁴ MV, TF, RL; Funding Acquisition: MV, JK

45 Abstract

46 Aphasia is a communication disorder that affects processing of language at different levels (e.g., acoustic,
47 phonological, semantic). Recording brain activity via EEG while people listen to a continuous story al-
48 lows to analyze brain responses to acoustic and linguistic properties of speech. When the neural activity
49 aligns with these speech properties, it is referred to as neural tracking. Even though measuring neural
50 tracking of speech may present an interesting approach to studying aphasia in an ecologically valid way,
51 it has not yet been investigated in individuals with stroke-induced aphasia. Here, we explored processing
52 of acoustic and linguistic speech representations in individuals with aphasia in the chronic phase after
53 stroke and age-matched healthy controls. We found decreased neural tracking of acoustic speech repres-
54 entations (envelope and envelope onsets) in individuals with aphasia. In addition, word surprisal displayed
55 decreased amplitudes in individuals with aphasia around 195 ms over frontal electrodes, although this
56 effect was not corrected for multiple comparisons. These results show that there is potential to capture
57 language processing impairments in individuals with aphasia by measuring neural tracking of continuous
58 speech. However, more research is needed to validate these results. Nonetheless, this exploratory study
59 shows that neural tracking of naturalistic, continuous speech presents a powerful approach to studying
60 aphasia.

61 **Keywords:** speech processing, neural tracking, aphasia, EEG, stroke

62 **Key points:**

- 63 • Individuals with aphasia display decreased encoding of acoustic speech properties (envelope and its
64 onsets) in comparison to healthy controls.
- 65 • Neural responses to word surprisal reveal decreased amplitudes in individuals with aphasia around
66 195 ms processing time (not corrected for multiple comparisons).
- 67 • Neural tracking of natural speech can be used to study speech processing impairments in aphasia.

68 1 Introduction

69 About one third of strokes result in aphasia, a language disorder that can impact auditory comprehension,
70 oral production, writing and/or reading (Engelter et al., 2006; National Aphasia Association, 2022; Pasley
71 and Knight, 2013). Aphasia can impact communication to different degrees, ranging from subtle to
72 severe impairments, and from recovery within hours after stroke to permanent language impairments.
73 The severity and persistence depend on factors such as lesion location and size, brain plasticity, therapy,
74 intrinsic motivation and social support (Pasley and Knight, 2013; Schevenels et al., 2020; Cordella et al.,
75 2022; Schevenels et al., 2022). In order to be effective, speech therapy should be given with high intensity
76 and the content should be tailored to the specific problems of each individual with aphasia (IWA) (Rohde
77 et al., 2018; Engelter et al., 2006; Schevenels et al., 2020; Brady, 2022). Individually tailored therapy
78 requires a precise diagnosis of language impairments.

79 Although diagnostic tests originally focused on the classic aphasia typology (i.e., assessing performance
80 on fluency, comprehension and repetition tasks), neuroimaging studies have suggested that focusing on
81 the different language processing components (i.e., acoustic, phonological, semantic, syntactic) is more in
82 line with the neural networks of language processing (e.g., the dual-stream model by Hickok and Poeppel
83 (2007)) (Rohde et al., 2018; Tremblay and Dick, 2016; Pasley and Knight, 2013; Wilson et al., 2023).
84 Fridriksson et al. (2018) suggest that lesions in the dorsal stream (i.e., sensori-motor integration) may
85 impair phonological processing, whereas damage to the ventral stream (i.e., acoustic-semantic integration)
86 may impair semantic processing. Hence, a neuroimaging approach focused on the linguistic aspects may
87 yield more precise diagnostic insights and a more effective therapeutic approach (Tremblay and Dick,
88 2016; Pasley and Knight, 2013). Moreover, behavioral testing after stroke is difficult because in 80% of
89 cases, IWA one year post-stroke have co-morbid cognitive problems, such as memory, executive functions
90 and/or attention problems, which can bias the results or even impede behavioral testing (El Hachioui
91 et al., 2014; Fonseca et al., 2018). Further, behavioral tests consist of artificial tasks that do not always
92 correspond to communication abilities in daily life. Thus, to provide targeted intervention and improve
93 recovery outcomes, an aphasia diagnosis that provides precise insights, that is less dependent on cognitive
94 performance and that is more ecologically valid is needed.

95 Electroencephalography (EEG) provides ways to study the brain's responses to speech with reduced active
96 participation of the patient. By averaging the EEG signal in response to a large number of repetitive
97 sound or speech stimuli, peaks in specific time ranges have been consistently identified in neurotypicals,
98 i.e., event-related potentials (ERPs), offering a window into the spatio-temporal patterns of the neural
99 response to speech. This way, ERPs related to acoustic (e.g., P1-N1-P2 complex (e.g., Martin et al.,
100 2008; Harris, 2020)) and linguistic (e.g., N400 (e.g., Hillyard and Kutas, 1984; Kutas and Federmeier,
101 2011; Nieuwland et al., 2020)) aspects of speech have been identified. In IWA, altered ERPs have been

102 found across language processing levels and across a variety of experimental stimuli and tasks (Ofek
103 et al., 2013; Becker and Reinvang, 2007; Ilvonen et al., 2001; Pulvermüller et al., 2004; Aerts et al., 2015;
104 Ilvonen et al., 2004; Pettigrew et al., 2005; Robson et al., 2017; Chang et al., 2016; Kawohl et al., 2010;
105 Khachatryan et al., 2017; Sheppard et al., 2017; Lice and Palmović, 2017; Kielar et al., 2012; Räling
106 et al., 2016). Most of these studies have reported decreased amplitudes and increased latencies in IWA
107 as compared to healthy controls, with the exception of the P2 in Aerts et al. (2015) and Ilvonen et al.
108 (2001), which observed opposite patterns. Some of these studies have found ERP amplitudes or latencies
109 to be correlated with language performance (Pettigrew et al., 2005; Robson et al., 2017; Khachatryan
110 et al., 2017).

111 The potential of ERPs to serve as evaluatory measure of intervention effects has recently been reviewed by
112 Cocquyt et al. (2020), who concluded that there is potential for ERPs to assess levels of aphasia symptoms,
113 after development of normative data. However, to date ERPs are not commonly used in the clinic. This
114 is likely due to small sample sizes and the heterogeneity of aphasia symptoms within the studied samples
115 (Silkes and Anjum, 2021), which complicates the development of validated norms for ERPs. While ERPs
116 are useful to understand the functional meaning of different peaks in the spatio-temporal patterns of the
117 neural response, their application in aphasia diagnostics poses further challenges, e.g., long administration
118 time due to different paradigms at distinct speech processing levels and the need for a large number of
119 repetitive stimuli to average across (Kandylaki and Bornkessel-Schlesewsky, 2019). Moreover, listening
120 to repetitive and artificially created stimuli is not representative of everyday language situations that
121 IWA struggle with mostly (Le et al., 2018). More naturalistic speech stimuli, such as a narrative, would
122 present a more ecologically valid stimulus to analyze the brain's response to speech (Lalor and Foxe,
123 2010; Ding and Simon, 2012; Hamilton and Huth, 2018; Kandylaki and Bornkessel-Schlesewsky, 2019;
124 Gillis et al., 2022).

125 From the narrative, different characteristics or representations of speech can be derived and their relation
126 to the EEG signal can be measured. When the neural signals align with the speech properties, it is referred
127 to as neural tracking. By examining the data in this way, spatial and temporal neural response properties
128 in response to multiple speech representation levels (e.g., acoustic, phonological, semantic, syntactic) can
129 be analyzed from the same data (Di Liberto et al., 2015; Brodbeck et al., 2018; Gillis et al., 2021, 2022;
130 Mesik et al., 2021). Moreover, a measure of the strength with which the different speech representations
131 are encoded in the EEG signal can be computed. Research has shown that even relatively short EEG
132 recordings (i.e., 10–20 minutes) can provide valid results with this approach (Di Liberto and Lalor, 2017).
133 Furthermore, limited active participation is required from the participant during such a paradigm, which
134 is especially advantageous for testing IWA. These characteristics make neural tracking an ideal tool to
135 study aphasia.

136 Examining neural tracking, the most frequently studied speech representation to date is the speech
137 envelope, consisting of the slow amplitude modulations of speech over time (Aiken and Picton, 2008). The
138 envelope presents an essential cue for speech intelligibility (Aiken and Picton, 2008; Shannon et al., 1995).
139 Whereas envelope tracking has not yet been investigated in individuals with stroke-induced aphasia, Dial
140 et al. (2021) have explored it in individuals with primary progressive aphasia (PPA). Individuals with
141 the logopenic variant of this neurodegenerative disease displayed increased envelope tracking compared
142 to healthy controls in the theta band (Dial et al., 2021). On the other hand, envelope tracking in other
143 disorders, such as developmental dyslexia, have shown decreased envelope tracking compared to controls
144 in the 1-8 Hz range, though the results were largely driven by the delta band (Di Liberto et al., 2018).
145 These studies show that neural tracking of continuous speech presents a promising new avenue to study
146 language disorders.

147 While envelope tracking is mostly considered an acoustic process in the literature, it has been found to
148 be affected by speech intelligibility and higher-level speech-specific processes (Peelle et al., 2013; Van-
149 thornhout et al., 2018; Broderick et al., 2019; Prinsloo and Lalor, 2022). This is not surprising given that
150 the speech envelope also encompasses important cues for segmentation (i.e., rise and fall times to iden-
151 tify acoustic edges) of the continuous speech signal into discrete units (i.e., phonemes, syllables, words,
152 phrases) (Aiken and Picton, 2008). Moreover, the syllable stress is also comprised in the envelope, which
153 gives the listener an indication of the prosody and thus even conveys linguistic information. Given this
154 entanglement of acoustic and linguistic cues, the speech envelope alone may not be ideal to find specific
155 neurophysiological correlates of acoustic, phonological and lexical semantic processing in aphasia.

156 Recently, neural tracking of higher-level linguistic speech representations has been investigated, i.e.,
157 speech representations containing information about phonemes and words that take into account linguistic
158 context (Broderick et al., 2018; Brodbeck et al., 2018; Weissbart et al., 2019; Gillis et al., 2021). This
159 way, it has been observed that older adults, in comparison to younger adults, have an altered neural
160 response to linguistic processes (Broderick et al., 2021; Mesik et al., 2021; Gillis and Kries et al., 2023).
161 To date, higher-level linguistic speech representations have not been investigated in IWA via neural
162 tracking, although higher-level speech processing impairments at the phonological and semantic level are
163 reported most frequently in IWA (in contrast to acoustic processing impairments).

164 In the present study, we investigated neural tracking of acoustic and linguistic speech representations
165 in individuals with post-stroke aphasia and healthy, age-matched controls. To this end, EEG data was
166 acquired while participants listened to a continuous story, of which 8 speech representations were derived.
167 The envelope and envelope onsets were considered to be acoustic representations of speech. The phoneme
168 and word onsets were considered representations of speech segmentation, i.e., at the interface between
169 acoustic and linguistic information. Lastly, phoneme surprisal and phoneme entropy were considered to

170 be linguistic representations at the phoneme level, while word surprisal and word frequency were regarded
171 as linguistic representations at the word level, i.e., related to lexical meaning. For the linguistic speech
172 representations, we controlled the variance explained by acoustic cues and vice-versa, as we aimed to
173 disentangle different levels of speech processing. Ultimately, disentangling mechanisms at different levels
174 of speech processing is necessary to investigate whether neural tracking will be useful as a diagnostic tool
175 for aphasia that provides information about different speech processing aspects. Here, our aim as a first
176 step towards this goal was to explore group differences between IWA and healthy controls based on neural
177 tracking of the different speech representations. Specifically, we studied group differences regarding the
178 strength of neural tracking, i.e., *how well* the brain tracks specific aspects of speech, and regarding *how*
179 the spatio-temporal pattern of the neural response operates during continuous speech perception. Based
180 on the aforementioned ERP and neural tracking studies, we expected to observe group differences in
181 speech representations at both acoustic and linguistic levels.

182 2 Materials & methods

183 2.1 Participants

184 We tested 41 IWA in the chronic phase after stroke (≥ 6 months) and 24 healthy controls that were
185 age-matched at group-level. 2 IWA had to be excluded post hoc - one because no lesion could be found
186 in the left hemisphere and one because we did not have access to any lesion information - resulting in a
187 sample size of 39 IWA. IWA were recruited in two ways. Between October 2018 and April 2022 (with a
188 COVID-19-related break between March and June 2020), patients were recruited via daily screening at
189 the stroke unit of the university hospital Leuven (score \leq cut-off threshold on the Language Screening
190 Test (LAST) (Flamand-Roze et al., 2011)) or via advertising the study in speech-language pathologists'
191 practices and rehabilitation centra (patients with a formal aphasia diagnosis) (see supplementary fig. S.1
192 for a detailed flowchart). Healthy age-matched controls were recruited via flyers positioned in recreational
193 community centers for elderly. The target age of healthy controls was gradually adapted based on the
194 mean age of IWA included in the study (supplementary figure S.2). The participants in this study are
195 partly overlapping with participants in Kries et al. (2023).

196 For data collection, we only included IWA that had no formal diagnosis of a psychiatric or neurodegen-
197 erative disorder and that had a left-hemispheric or bilateral lesion. All aphasia participants were tested
198 in the chronic phase after stroke (time since stroke onset in months (median(range)): 16.1(6-126.1). The
199 aphasia sample was checked for language impairments at the moment of data collection using two stan-
200 dardized diagnostic aphasia tests, i.e., the diagnostic test ScreeLing Visch-Brink et al. (2010) and the
201 Dutch picture-naming test (Nederlandse Benoemtest (NBT); Van Ewijk et al. (2020), using the same
202 procedure as reported in Kries et al. (2022, biorxiv). The ScreeLing was administered on a tablet using
203 the Gorilla Experiment Builder (<http://www.gorilla.sc>) (Anwyl-Irvine et al., 2020). The average score
204 by group is reported in table 1. We included individuals that scored either (1) below the cut-off threshold
205 (ScreeLing threshold: 68/72 points; NBT threshold: 255/276 points) on at least one of these two tests
206 at the moment of data collection (n=27) (supplementary table S.1; supplementary fig. S.3), or (2) had a
207 documented language impairment in the acute phase (n=12). Note that 10 out of the latter 12 IWA still
208 followed speech-language therapy at the time of data collection (supplementary table S.1).

209 All participants were Dutch native speakers from Flanders, Belgium. Informed consent was obtained
210 from all participants for the recruitment via screening and for the data collection in the chronic phase.
211 The study received ethical approval by the medical ethical committee of KU Leuven and UZ Leuven
212 (S60007) and is in accordance with the declaration of Helsinki.

213 In table 1, we summarized demographic information by group (details can be found in supplementary
214 table S.1). Age, sex, education, handedness and multilingualism did not differ between groups (age: W

215 = 464, p = 0.96; sex: $\chi^2 = 0$, df = 1, p = 1; education: $\chi^2 = 7.26$, df = 4, p = 0.1; handedness: χ^2
 216 = 0.063, df = 2, p = 0.98; multilingual: $\chi^2 = 0.182$, df = 1, p = 0.66). Details about the stroke in
 217 IWA, i.e., time since stroke onset, stroke type, occluded blood vessel, lesion location and speech-language
 218 therapy, can be found in supplementary table S.1. To visualize the damaged brain tissue of IWA, a lesion
 219 overlap image was created (fig. 1). More information on the lesion delineation process can be found in the
 220 supplementary material S.1.1.4. Demographic information was acquired via a self-reported questionnaire.
 221 Handedness was assessed via the Edinburgh Handedness Inventory (Oldfield, 1971).

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

| | | Control | Aphasia | |
|--|----------------------------|-------------|-------------|-------|
| Demographics | | | | |
| Age in years (mean(SD)) [◊] | | 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) | ** |

[◊] Age is also used as a covariate; n.s. = not significantly different; p<0.05*; p<0.01**; p<0.001***

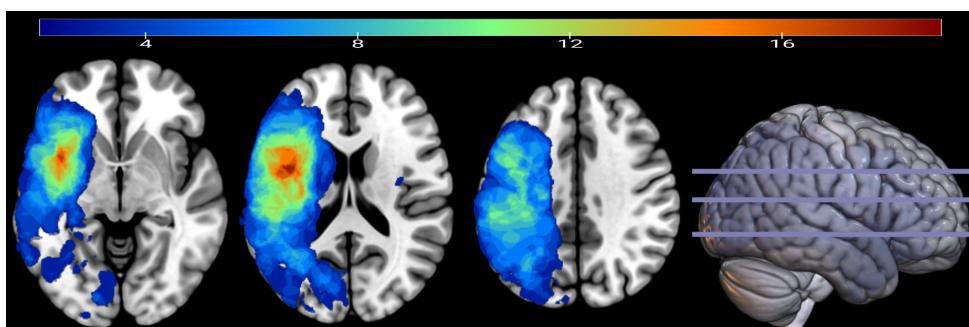


Figure 1: **Lesion overlap image of the aphasia sample.** The maximum overlap corresponds to 19 out of the total sample of 39 individuals with aphasia. Axial slices are shown in neurological orientation.

222 2.2 Behavioral measures that serve as covariates

223 2.2.1 Hearing

224 Hearing thresholds were assessed via pure tone audiometry (PTA) at frequencies ranging from 0.25 to 4
225 kHz. In case the hearing thresholds below 4 kHz were >25 dB HL, this information was used to increase
226 the amplitude of stimulus presentation during the EEG measurement. The PTA thresholds at 0.25, 0.5
227 and 1 kHz were averaged and then divided in half to come to the amount of dB that was added to the
228 stimulus presentation amplitude of 60 dB SPL during the EEG paradigm. This calculation was done for
229 each ear separately. After a short example stimulus, participants were asked whether the loudness was
230 comfortable and if necessary the presentation volume was adjusted. The degree of volume adjustment did
231 not differ between groups ($W = 170$, $p = 0.8$). Furthermore, hearing thresholds were used as covariates
232 in statistical models (section 2.4). For this purpose, the Fletcher index (average of hearing thresholds at
233 0.5, 1 and 2 kHz) was calculated per ear and subsequently averaged across both ears. The Fletcher index
234 did not differ between IWA and healthy controls ($W = 541.5$, $p = 0.29$, confidence interval: [-2.49 10];
235 table 1).

236 2.2.2 Cognition

237 The Oxford Cognitive Screen-NL was administered to assess cognitive functioning (Huygelier et al., 2019).
238 This test was designed to be language-independent, such that cognitive functioning can be disentangled
239 from language functioning, which is especially important for IWA. Due to limited time in the testing
240 protocol, we chose to only assess 4/10 subscales, i.e. attention and hemispatial neglect, reading, executive
241 functioning, and memory. Hemispatial neglect was used as a means to potentially exclude participants
242 in case they had too severe hemineglect, which could bias outcomes at most of the administered tests.
243 However, the highest hemineglect score was still at a very mild level and thus we decided to not exclude
244 any participants based on hemispatial neglect.

245 The task to assess attention consisted of crossing out target shapes among distractor shapes. The task
246 to assess executive functions consisted of connecting circles and triangles in alternation in descending
247 order of size. The memory task consisted of free recall and recognition of words (from the sentence read
248 for the reading task) and shapes. These 3 tasks were used to calculate a composite score of cognitive
249 functioning. This score was calculated by transforming the raw scores of each test into percentages and
250 then averaging across the three outcomes. The composite score was used to regress out differences in
251 cognitive functioning to explore neural tracking differences between groups. The cognition composite
252 score was significantly lower in IWA than in healthy controls ($W = 209.5$, $p < 0.001$).

253 **2.2.3 Alertness and fatigue**

254 Given that the experimental protocol (behavioral and EEG testing) was relatively long, especially con-
255 sidering that IWA often have cognitive impairments (e.g., attention), we decided to monitor the alertness
256 and tiredness or fatigue at 3 time points throughout the experimental protocol (referred to as t1, t2
257 and t3). In supplementary figure S.4 E, the experimental protocol with reference to the timing of the
258 alertness and fatigue questions is visualized. t1 was administered right at the start of the testing session,
259 t2 after the EEG measurement and t3 at the end of the experimental protocol. Participants had to
260 indicate on a Likert scale of 1 to 5 how alert and how tired they were at that moment. The questions
261 were presented visually and auditory at the same time, as visualized in supplementary figure S.4 A and
262 B. In supplementary section S.4, we describe the results of the interaction analysis between group and
263 time points (supplementary fig. S.4 C and D, section S.1.1.5). Given that neural tracking has been shown
264 to be influenced by attention (Lesenfants and Francart, 2020), we used the average ratings of t1 and t2
265 of the alertness and fatigue scale respectively as covariates in the analysis concerning group differences
266 in neural tracking of speech. The average of t1 and t2 scores was specifically used because the EEG
267 measurement took place in between these moments. A group difference was found for alertness ($W =$
268 243, $p = 0.002$) and for fatigue ($W = 209$, $p = 0.001$). IWA were on average less alert and more tired
269 than healthy controls at t1 and t2 combined.

270 **2.3 EEG-based measures**

271 **2.3.1 Experimental paradigm**

272 The EEG measurements took place in a soundproof room with Faraday cage. We recorded 64-channel
273 EEG (ActiveTwo, BioSemi, Amsterdam, NL) at a sampling frequency of 8192 Hz. Participants were
274 instructed to listen to a 24 minute long story while EEG data was recorded. They were seated one meter
275 away from a screen and were asked to look at a fixation cross while listening in order to minimize eye
276 movement artifacts in the EEG signal (fig. 2 A). The story *De wilde zwanen* (The Wild Swans), written
277 by Hans Christian Andersen and narrated by a female Flemish-native speaker, was cut into 5 parts of
278 on average 4.84 minutes (standard deviation (SD): 9.58 seconds) each. The silences in the story were
279 reduced to 200 ms duration and the sample rate was set to 48 kHz. The software APEX (Francart et al.,
280 2008) was used to calibrate and present stimuli. The story was presented bilaterally via shielded ER-3A
281 insert earphones (Etymotic Research) at an amplitude of 60 dB SPL (A weighted), except if hearing
282 thresholds were above 25 dB HL at the PTA, in which case the presentation volume was augmented (see
283 section 2.2.1).

284 After each story part, participants answered a yes/no question and a multiple choice question about the
285 content of the preceding story part. As these questions were not validated, we did not assess them. They

286 were solely introduced in the protocol to make participants follow the content of the story attentively.
287 Nonetheless, 5 participants (4 IWA, 1 control) fell asleep during parts of the EEG measurement. Given
288 that an awake state is necessary to follow the contents of a story (such as reflected in linguistic speech
289 representations) (Makov et al., 2017), we decided to exclude these story parts from the analysis. For
290 one other control participant, a part of the data was not saved correctly and could thus also not be
291 used for analysis. This means that for 6 participants, less than 24 minutes of data was used for analysis
292 (19.36 minutes of data for 1 control and 1 aphasia participant, 14.52 minutes for 1 control and 2 aphasia
293 participants, 9.68 minutes for 1 aphasia participant).

294 2.3.2 EEG signal processing

295 The EEG signal processing was performed in MATLAB (version 9.1.0.441655 (R2016b)). The EEG data
296 of the 5 story parts (i.e., epochs) were concatenated. Eye movement artifact removal was implemented
297 using a multichannel Wiener filter (Somers et al., 2018). The EEG signal was referenced to the common
298 average. For high-pass filtering, a least squares filter was applied with a filter order of 2000, with a
299 passband frequency of 0.5 Hz and a stopband frequency of 0.45 Hz. For low-pass filtering, a least squares
300 filter with a filter order of 2000 was applied with a passband frequency of 25 Hz and a stopband frequency
301 of 27.5 Hz. The EEG data was downsampled to 128 Hz and subsequently normalized by subtracting the
302 mean and dividing by the SD per epoch.

303 2.3.3 Neural tracking

304 To investigate neural tracking, we used a forward modeling approach (i.e., encoding model), meaning
305 that speech representations were used to predict the EEG signal (Mesgarani et al., 2014; Di Liberto
306 et al., 2015; Holdgraf et al., 2017). Here, we were interested in both acoustic and linguistic speech
307 representations. We relied on 8 representations, which were extracted from the stimulus story (fig. 2 B).
308 We considered the envelope and envelope onsets as acoustic speech representations. Phoneme and word
309 onsets represented phoneme- and word-level segmentation of speech. Phoneme surprisal and phoneme
310 entropy were considered as linguistic representations at the phoneme level, word surprisal and word
311 frequency at the word level.

312 **Envelope.** The envelope was extracted by using a gammatone filter bank of 28 channels with center
313 frequencies between 50 and 5000 Hz. We applied a power law on the absolute values and averaged
314 across the 28 envelopes (same parameters as in Vanthornhout et al. (2018)). These steps were
315 applied because they model the auditory system's structure (Biesmans et al., 2017).

316 **Envelope onsets.** The envelope onsets were calculated as the half-wave rectification of the first deriva-
317 tive of the envelope.

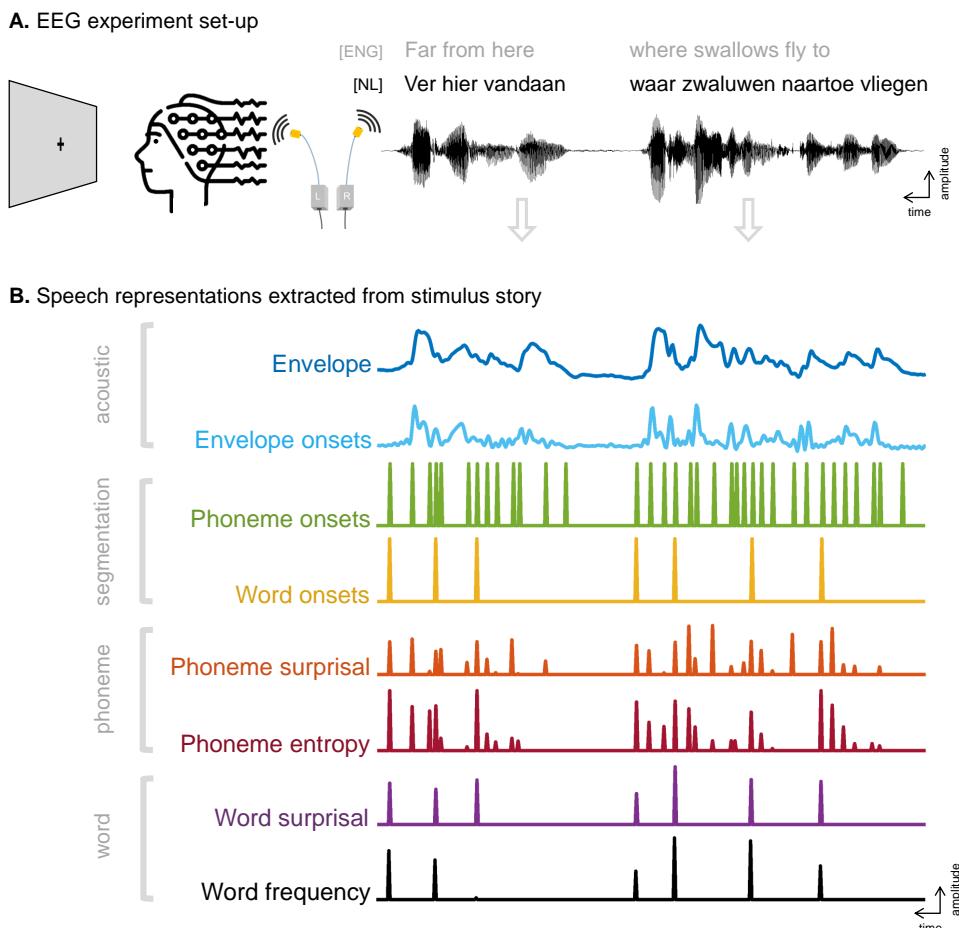


Figure 2: **EEG experiment set-up and extraction of speech representations from the stimulus story.** **A.** Participants listened to a story in Dutch (Flemish dialect) while EEG data was recorded. They were asked to look at a fixation cross while listening. The first 2 phrases of the story are visualized as written text (with an English translation) and audio signal. **B.** From the audio signal as depicted in panel A, 8 speech representations were extracted that reflect acoustic and linguistic properties of the story.

318 **Phoneme and word onsets.** Phoneme and word onsets were coded as dummy variables with a pulse
319 at the beginning of each phoneme, respectively of each word. In order to get there, an aligner
320 (Duchateau et al., 2009) was used to create alignment files containing the timing of each phoneme,
321 respectively each word for the audio files of the stimulus story.

322 **Phoneme surprisal.** Phoneme surprisal was computed as the negative logarithm of the phoneme prob-
323 ability in the activated cohort. The activated cohort refers to words activated by initial phonemes,
324 e.g., after hearing the sound /pl/, the activated cohort consists of words such as play, plus and
325 plural. Phoneme surprisal is thus a representation of how surprising a phoneme is given the acti-
326 vated cohort. The first phoneme of each word included all words in the active cohort. Phoneme
327 surprisal was calculated based on the SUBTLEX-NL database (Keuleers et al., 2010) and a custom
328 pronunciation dictionary.

329 **Phoneme entropy.** Phoneme entropy is a measure of the degree of competition between the words
330 congruent with the current phonemic input. For instance, after hearing the sounds /pl/, many
331 possible words are present in the activated cohort ($n=999$), mirrored in a high degree of competition.
332 Yet, after the next phonemes, the number of possible words decreases (e.g., for /plu/, $n=162$ and for
333 /plur/, $n=19$), and thus the activated cohort decreases, reflected in a lower degree of competition.
334 The degree of competition was computed as the Shannon entropy of the words in the activated
335 cohort. Again, the first phoneme of each word included all words in the active cohort. Phoneme
336 entropy was also calculated based on the SUBTLEX-NL database (Keuleers et al., 2010) and a
337 custom pronunciation dictionary.

338 **Word surprisal.** Word surprisal was calculated as the negative logarithm of the conditional probability
339 of a given word based on the 4 previous words. Word surprisal thus represents how surprising a
340 word is, taking into account the 4 previous words. Word surprisal was calculated using the 5-gram
341 model by Verwimp et al. (2019).

342 **Word frequency.** Word frequency was calculated as the negative logarithm of the unigram probability
343 and represents how frequently words are used. Given that we used the negative logarithm, words
344 with a higher frequency are reflected in lower scores. Word frequency was also calculated using the
345 5-gram model by Verwimp et al. (2019).

346 **Isolating speech processing levels.** An issue when analyzing acoustic and linguistic speech represen-
347 tations is their collinearity, e.g., some top-down linguistic cues at the phoneme and word level are also
348 represented in the speech envelope, such as word boundaries and syllable stress, which contain semantic
349 cues. Vice versa, some bottom-up acoustic information is also present in linguistic phoneme and word
350 level representations (Gillis et al., 2022), e.g., due to amplitude rises and falls defining the boundaries

351 between (pre-)lexical units. Neural signals related to acoustic processing can thus also be captured by
352 neural tracking of linguistic speech representations and vice versa.

353 In this study, we were interested in disentangling - as far as possible - different speech processing mech-
354 anisms. This is important should encoding/decoding modeling be used for diagnosing different language
355 profiles of aphasia in the future, e.g., disentangle individuals that have more problems with acoustic,
356 phonological or semantic processing. Therefore, we regressed out the variance explained by the speech
357 representations that we were not interested in. Specifically, we first applied an ordinary least squares
358 (OLS) regression analysis without regularization, using the EEG signal as dependent variable and speech
359 representations as predictors that share variance with the representations of interest (table 2). All the
360 data was used for training and testing, such that all activity related to the collinear speech representations
361 was regressed out. As a second step, we then used the residual EEG signal of that regression as input for
362 the encoding model in order to model the relationship with the speech representations of interest. The
363 constellation of speech representations used as predictors in the encoding models is illustrated in table 2.

Table 2: Constellation of speech representations in the 4 encoding models.

| Regressing out representations not of interest | Boosting on residuals (representations of interest) | Encoding model/ referred to as |
|---|---|-----------------------------------|
| phoneme onsets word onsets phoneme surprisal phoneme entropy word surprisal word frequency | envelope envelope onsets | acoustic |
| envelope envelope onsets phoneme surprisal phoneme entropy word surprisal word frequency | phoneme onsets | phoneme-level segmentation |
| envelope envelope onsets phoneme surprisal phoneme entropy word surprisal word frequency | word onsets | word-level segmentation |
| envelope envelope onsets phoneme onsets word onsets* | phoneme surprisal phoneme entropy | phoneme-level linguistic |
| envelope envelope onsets phoneme onsets word onsets* | word surprisal word frequency | word-level linguistic |

*We decided against regressing out phoneme level linguistic representations for the word level model and vice-versa, because we expected that more data would be needed to get meaningful results from such an analysis.

364 **Computation of temporal response function and prediction accuracy.** For the encoding analysis
365 of the speech representation of interest, we applied the boosting procedure (David et al., 2007). We made
366 use of the Eelbrain toolbox for this step (<https://doi.org/10.5281/zenodo.6992921>) (Brodbeck et al.,
367 2021). The data was split into a held-out test set of roughly 2 minutes of the data and a training set
368 containing the rest of the data. The training set was used to perform regression analyses on the residual

369 EEG signal with the speech representation as predictor, for a number of time-shifted versions, resulting
370 in the temporal response function (TRF). The number of time-shifted versions was defined by the chosen
371 integration window length, i.e., -0.1078 to 0.6109 seconds, and the sampling frequency ($((\text{ending time}$
372 $\text{lag-starting time lag})/(1/\text{sampling frequency}))$), which resulted in 92 time shifts. The TRF thus provides
373 information about *how* the neural response pattern operates across processing time. This information
374 and the speech representation were then used to predict EEG signals for the test set of the data. The
375 predicted EEG signal was correlated with the originally recorded EEG signal, providing a measure of
376 *how well* the speech representation is encoded in the brain for each electrode, hereafter referred to as
377 prediction accuracy. The higher the prediction accuracy is, the stronger the speech representation is
378 encoded in the EEG signal. This procedure was repeated for different partitions of the data into training
379 and test sets, such that each part of the data was once used as test set, resulting in 12 folds (i.e., k-fold
380 cross-validation). For the TRF, the 12 folds were averaged across to get robust outcome measures. To
381 arrive at the final prediction accuracy, the folds of the predicted EEG signal were concatenated and
382 subsequently correlated with the originally recorded EEG signal.

383 2.3.4 Determination of the TRF peak latency ranges

384 To determine in which time windows to perform cluster-based permutation tests, we used the built-in
385 in MATLAB function *findpeaks* (MATLAB, 2016) to identify peaks in the TRF. Peak latencies were
386 extracted per speech representation and per participant. An arbitrary range of 150 ms was defined
387 around the average latency of the control group per identified peak (table 3). The average latency per
388 peak of the aphasia group was very similar to the control group (the largest difference among all peaks
389 was 18 ms), such that the selected ranges were valid to find differences between the aphasia and control
390 group. For time ranges that started before 0, the lower bound was set to 0. The 150 ms ranges, as
391 indicated in table 3 and visualized in supplementary figure S.5, were used as time windows to perform
392 cluster-based permutation tests on the TRF (section 2.4 for details). As 26 peaks were found across
393 speech representations, we conducted 26 cluster-based permutation tests.

394 2.4 Statistical analysis

395 Statistical analyses were performed in R (R Core Team, 2017) and in the Python (Van Rossum and
396 Drake Jr, 1995) toolbox Eelbrain (Brodbeck, 2020).

397 2.4.1 Group comparison of the strength of neural tracking

398 For the prediction accuracy analysis, we averaged across all 64 electrodes and conducted a linear model
399 in R in order to investigate group differences and to regress out the covariates, i.e., *average prediction*
400 *accuracy* ~ *group* + *age* + *hearing* + *cognition* + *alertness* + *fatigue*. We repeated this for each of

Table 3: TRF peak ranges of the control group used for defining time windows for cluster-based permutation testing.

| Speech representation | Average peak latency (ms) | Range (ms) |
|-----------------------|---------------------------|------------|
| envelope | 29 | 0 - 104 |
| | 163 | 88 - 238 |
| envelope onsets | 44 | 0 - 119 |
| | 102 | 27 - 177 |
| | 209 | 134 - 284 |
| phoneme onsets | 61 | 0 - 136 |
| | 252 | 177 - 327 |
| | 305 | 230 - 380 |
| word onsets | 114 | 39 - 189 |
| | 190 | 115 - 265 |
| | 308 | 233 - 383 |
| | 446 | 371 - 521 |
| phoneme surprisal | 221 | 146 - 296 |
| | 360 | 285 - 435 |
| | 354 | 279 - 429 |
| phoneme entropy | 243 | 168 - 318 |
| | 389 | 314 - 464 |
| | 416 | 341 - 491 |
| word surprisal | 122 | 47 - 197 |
| | 199 | 124 - 274 |
| | 320 | 245 - 395 |
| | 428 | 353 - 503 |
| word frequency | 122 | 47 - 197 |
| | 197 | 122 - 272 |
| | 330 | 255 - 405 |
| | 459 | 384 - 534 |

401 the 4 encoding models, i.e. the acoustic, phoneme onsets, word onsets, phoneme and word level model.

402 We checked the normality assumptions using the Shapiro-Wilk test, which failed to reject H0 for each
403 model. The homogeneity of variances assumption was not met in any of the 5 models. Nonetheless, we
404 interpreted the linear models, given that the residuals were normally distributed. As each model was
405 based on a different dataset, we did not control for multiple comparisons in this analysis.

406 As a second analysis of the prediction accuracies, we conducted a cluster-based permutation test to see if
407 certain electrodes drive the difference between groups. To this end, we used the function *testnd.TTestIndependent*
408 from the Eelbrain toolbox (<https://doi.org/10.5281/zenodo.6992921>). The cluster-based permutation test
409 is a mass-univariate independent samples t-test that relies on bootstrapping (see Gillis et al. (2021) for
410 more details). We defined a maximum p-value of 0.05 as threshold. We report the number of electrodes
411 in the significant cluster, the v-value and the p-value.

412 **2.4.2 Group comparison of the neural response pattern**

413 To investigate the TRF pattern differences between the control group and the aphasia group, we applied
414 cluster-based permutation tests in the arbitrary integration window ranges identified by the peak latency
415 extraction (table 3). The same function and parameters were used as for the prediction accuracy. As
416 there were multiple peaks identified for each of the 8 speech representations, we corrected the p-value for
417 multiple comparisons within speech representation using the false discovery rate (FDR). Given that this

⁴¹⁸ is an exploratory study, we report the uncorrected and corrected significance thresholds.

419 3 Results

420 3.1 Response accuracy to the questions asked during the EEG paradigm

421 Participants listened to 5 story parts of ca. 5 minutes each while EEG data was recorded. After each
422 part, a yes/no and a multiple choice question were asked about the preceding story part. Figure 3 shows
423 the response accuracy per group separately for the yes/no question and the multiple choice question.
424 For both types of questions, a significant group difference was found (Yes/No question: $W = 300$, $p =$
425 0.01; Multiple choice question: $W = 180$, $p < 0.0001$). These questions were however not validated and
426 therefore, this result should not be interpreted, but instead be seen as descriptive.

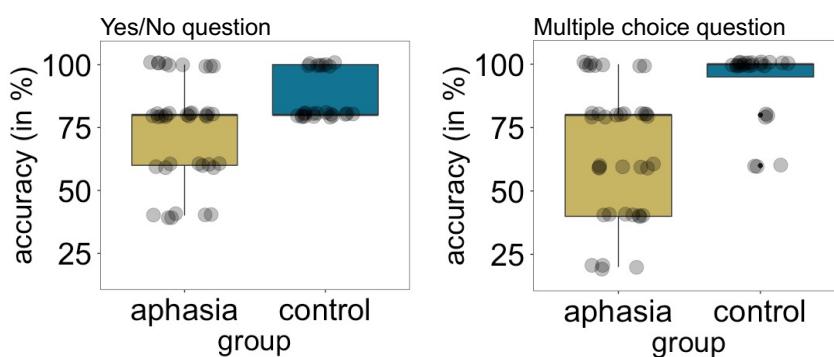


Figure 3: **Response accuracy by group for the yes/no questions and the multiple choice questions** after each of the 5 story parts that served as stimuli during the EEG measurement. For both types of questions, a significant group difference was found.

427 3.2 Prediction accuracy

428 When averaging the prediction accuracy across all electrodes, we found a significant group difference
429 in the acoustic model (group effect: $F = 7.11$, $p = 0.01$) (fig. 4A). The healthy control group showed
430 higher prediction accuracies in comparison to the aphasia group. The group effect was present despite
431 controlling for the influence of age, hearing, cognition, alertness, fatigue and lesion size in the model.
432 None of the covariates significantly explained the prediction accuracies, except for the acoustic model.
433 Neither the phoneme onset and word onset models nor the phoneme- and word-level linguistic models
434 showed a significant group difference. The results of all models are reported in table 5. Additionally, we
435 also report the results when the covariates are not included in the model (table 4).

436 In order to get an idea how the lesion size affects the link between neural tracking measures and group,
437 we created a hypothetical, statistical scenario in which IWA would have a lesion size of 0. We did this
438 by comparing the intercept of the linear model *average prediction accuracy ~ lesion size* including IWA
439 only, to the intercept of the model *average prediction accuracy ~ group* including both IWA and controls.
440 We found that the 2 intercepts are not significantly different for any of the speech representation models
441 (acoustic: $p=0.71$; phoneme onsets: $p=0.51$; word onsets: $p=0.80$; phoneme linguistic: $p=0.72$; word

Table 4: Group comparison results on the prediction accuracy models (without covariates).

| Effect | adj.R ² | Df | F | Estimate | Standard error | P |
|----------------------------------|--------------------|--------|------|----------|----------------|--------------|
| Acoustic | | | | | | |
| Group | 0.103 | (1,61) | 8.12 | -0.00 | 0.002 | 0.006 |
| Phoneme onsets | | | | | | |
| Group | 0.032 | (1,61) | 3.10 | -0.00 | 0.00 | 0.083 |
| Word onsets | | | | | | |
| Group | 0.005 | (1,61) | 1.34 | -0.00 | 0.00 | 0.251 |
| Phoneme-level linguistics | | | | | | |
| Group | 0.014 | (1,61) | 1.94 | -0.00 | 0.00 | 0.168 |
| Word-level linguistics | | | | | | |
| Group | 0.02 | (1,61) | 2.27 | -0.00 | 0.00 | 0.137 |

Table 5: Results of the group effect and covariates on the prediction accuracy models.

| Effect | adj.R ² | DF | F | Estimate | Standard error | p |
|----------------------------------|--------------------|---------|-------|----------|----------------|-------------|
| Acoustic | | | | | | |
| Model performance | 0.083 | (6, 52) | 1.885 | | 0.013 | 0.1 |
| Group | | 1 | 7.11 | -0.00 | 0.00 | 0.01 |
| Age | | 1 | 0.16 | 0.00 | 0.00 | 0.69 |
| Hearing | | 1 | 0.05 | 0.00 | 0.00 | 0.82 |
| Cognition | | 1 | 0.16 | 0.00 | 0.00 | 0.69 |
| Alertness | | 1 | 3.58 | 0.00 | 0.00 | 0.06 |
| Fatigue | | 1 | 0.25 | 0.00 | 0.00 | 0.61 |
| Phoneme onsets | | | | | | |
| Model performance | 0.047 | (6, 52) | 1.483 | | 0.004 | 0.2 |
| Group | | 1 | 3.06 | -0.00 | 0.00 | 0.08 |
| Age | | 1 | 0.69 | -0.00 | 0.00 | 0.40 |
| Hearing | | 1 | 3.09 | 0.00 | 0.00 | 0.08 |
| Cognition | | 1 | 0.69 | 0.00 | 0.00 | 0.40 |
| Alertness | | 1 | 0.12 | 0.00 | 0.00 | 0.73 |
| Fatigue | | 1 | 1.25 | -0.00 | 0.00 | 0.26 |
| Word onsets | | | | | | |
| Model performance | -0.038 | (6, 52) | 0.638 | | 0.004 | 0.69 |
| Group | | 1 | 1.58 | -0.00 | 0.00 | 0.21 |
| Age | | 1 | 0.28 | -0.00 | 0.00 | 0.59 |
| Hearing | | 1 | 0.94 | 0.00 | 0.00 | 0.33 |
| Cognition | | 1 | 0.26 | -0.00 | 0.00 | 0.61 |
| Alertness | | 1 | 0.69 | 0.00 | 0.00 | 0.41 |
| Fatigue | | 1 | 0.08 | -0.00 | 0.00 | 0.77 |
| Phoneme-level linguistics | | | | | | |
| Model performance | 0.002 | (6, 52) | 1.024 | | 0.005 | 0.42 |
| Group | | 1 | 2.60 | -0.00 | 0.00 | 0.11 |
| Age | | 1 | 0.00 | -0.00 | 0.00 | 0.97 |
| Hearing | | 1 | 1.79 | 0.00 | 0.00 | 0.18 |
| Cognition | | 1 | 0.53 | -0.00 | 0.00 | 0.47 |
| Alertness | | 1 | 0.36 | 0.00 | 0.00 | 0.55 |
| Fatigue | | 1 | 0.86 | -0.00 | 0.00 | 0.3568 |
| Word-level linguistics | | | | | | |
| Model performance | -0.01 | (6, 52) | 0.903 | | 0.004 | 0.49 |
| Group | | 1 | 2.46 | -0.00 | 0.00 | 0.12 |
| Age | | 1 | 0.46 | 0.00 | 0.00 | 0.49 |
| Hearing | | 1 | 1.40 | 0.00 | 0.00 | 0.24 |
| Cognition | | 1 | 0.15 | -0.00 | 0.00 | 0.70 |
| Alertness | | 1 | 0.26 | 0.00 | 0.00 | 0.61 |
| Fatigue | | 1 | 0.69 | -0.00 | 0.00 | 0.40 |

DF=degrees of freedom; significant effects are marked in bold.

442 linguistic: $p=0.64$). This means that, if IWA had no lesion (a lesion size of 0), then they would not differ
443 from healthy controls on measures of neural tracking.

444 The slope coefficient of the model *average prediction accuracy ~ lesion size* including IWA only was not
445 significantly different from 0 for the prediction accuracy models acoustics, word onsets, phoneme-level
446 linguistics and word-level linguistics (acoustic: $p=0.08$; word onsets: $p=0.59$; phoneme linguistic: $p=0.13$;
447 word linguistic: $p=0.59$). The phoneme onset model however showed a slope coefficient of lesion size
448 that was significantly different from 0 (phoneme onsets: estimate=-2.35e-05, $p=0.01$), meaning that
449 a larger lesion size is associated with lower neural tracking scores within the aphasia group, for the
450 phoneme onset model. A relation between lesion size and the phoneme onsets model was also found in a
451 correlation analysis within the aphasia group (phoneme onsets model: Spearman's $r = -0.355$, $p = 0.026$;
452 supplementary fig. S.6).

453 The cluster-based permutation tests revealed a group difference for the acoustic model, and additionally
454 provided information about what cluster of electrodes is driving the difference between the control group
455 and the aphasia group. Specifically, 2 clusters were found to differ between groups (fig. 4 B). A frontal
456 left-lateralized cluster (number of sensors = 9, $v = 25.012$, $p = 0.02$) and a posterior cluster (number
457 of sensors = 15, $v = 44.516$, $p = 0.004$) both displayed decreased prediction accuracies in IWA. The
458 phoneme onsets model revealed one cluster that differed between groups (number of sensors = 6, $v =$
459 15.511, $p = 0.03$), namely displaying decreased prediction accuracies in IWA. Neither the word onset
460 model nor the phoneme- and word-level linguistic models showed any significant group differences.

461 Aphasia is a disorder with a heterogeneous phenotype, as was reflected in the large variability of the
462 language test outcomes (see supplementary table S.1). This made us wonder whether the heterogeneity
463 of language levels in the aphasia group may hide subtle effects in group statistics. Therefore, we decided
464 post-hoc to split up the aphasia group into a more mild and a more severe aphasia group. In the
465 supplementary material (S.1.4), we repeated the group comparison analyses after splitting up the aphasia
466 group. The criteria for splitting up the group is described in more detail in the supplementary material
467 (also see supplementary fig. S.7 and S.10). When averaging the prediction accuracy across all electrodes,
468 we found that the group difference for the acoustic model was not present for the comparison of the
469 control and mild aphasia subgroup, but was present for the comparison of the control group and the
470 aphasia subgroup with more severe language difficulties (supplementary fig. S.8 A). No group effects
471 were observed for any of the other 4 models. The cluster-based permutation test of the acoustic model
472 showed 2 almost identical clusters to the ones found in the analyses with 2 groups (fig. 4 B), however
473 only for the comparison between the control group and the aphasia subgroup with more severe language
474 difficulties (supplementary fig. S.8 B).

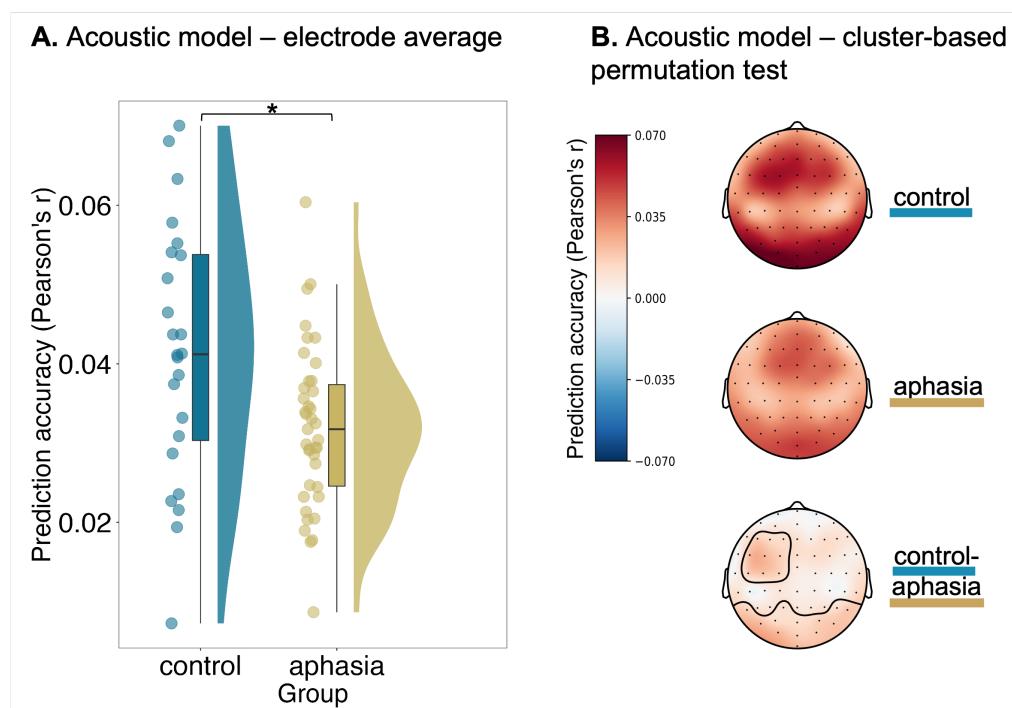


Figure 4: **Individuals with aphasia display decreased neural tracking of acoustic speech representations - across all electrodes and in local clusters.** **A.** When averaging the prediction accuracies of the acoustic model across all 64 electrodes, we found a significant group difference, even when we controlled for age, hearing, cognition, alertness and fatigue. The acoustic model consists of the speech envelope and its onsets as speech representations. **B.** The cluster-based permutation test revealed 2 clusters that significantly differed between groups, showing lower prediction accuracies in individuals with aphasia. The lowest topplot consists of the difference between the control group and the aphasia group and the significant clusters are contoured.

475 3.3 Temporal response function

476 For the speech envelope TRF, we found 2 clusters that differed between the control and aphasia group
477 around 180 ms (fig. 5 A; frontal cluster: time(ms) = [95 243], number of sensors = 28, v = 561.96, p= 0.001; posterior cluster: time(ms) = [142 243], number of sensors = 18, v = -425.69, p = 0.005). The 2
478 clusters are similar to the clusters from the prediction accuracy of the acoustic model (fig. 4 B). After
479 correction for multiple comparisons via FDR (for n=2 comparisons due to the tested time ranges per
480 speech representations, see table 3), these results remained significant (frontal left: p = 0.002; posterior:
481 p = 0.01). Looking at the neural response to phoneme onsets, we found a cluster that differed between
482 the control group and the aphasia group over frontal electrodes around 280 ms (time(ms) = [235 384],
483 number of sensors = 11, v = 240.58, p = 0.04). This group effect in the neural response to phoneme
484 onsets did not survive correction for multiple comparisons.

485
486 No group difference clusters were found for the speech representations envelope onsets, phoneme surprisal
487 and phoneme entropy. However, for the speech representations at the word level, i.e. word onsets and
488 word surprisal, displayed clusters that differed between groups were observed (fig. 5 C and D). The neural
489 response patterns to word onsets displayed 2 clusters that differed between the control and aphasia group
490 around 195 ms (frontal left-lateralized cluster: time(ms) = [134 243], number of sensors = 10, v = -220.3,
491 p = 0.03; posterior cluster: time(ms) = [118 267], number of sensors = 14, v = 267.25, p = 0.016).
492 The neural response patterns to word surprisal displayed a cluster that differed between the control and
493 aphasia group that spread across 2 negative peaks, around 135 ms (difference in posterior left electrodes)
494 and around 195 ms (difference in frontal left electrodes) (time(ms) = [56 204], number of sensors = 21,
495 v = -227.94, p = 0.01). None of these effects survived the correction for multiple comparisons via FDR.

496 Same as for the prediction accuracy, we split up the aphasia group into a mild aphasia subgroup and
497 an aphasia subgroup with more severe language difficulties and repeated these analyses for exploratory
498 purposes in the supplementary material (supplementary fig. S.9). For the envelope TRF, we found that
499 the control group and aphasia subgroup with more severe language impairments, but not the control and
500 mild aphasia subgroup, showed a significantly different neural response pattern in 2 similar clusters as
501 were found in the group comparison with the full aphasia group (supplementary fig. S.9). This difference
502 in clusters occurred as well around 180 ms, same as in the comparison of the control and full aphasia group.
503 For the envelope onsets TRF, a group difference was found around 200 ms between the control group and
504 the more severe aphasia subgroup. Moreover, the word-level representations also showed significantly
505 different clusters between the control group and more severe aphasia subgroup, but not between the
506 control and mild aphasia subgroup. However, these differences in clusters were found in a later time
507 window than the clusters found in the group comparison between the control and the full aphasia group
508 (200 ms), namely around 360-370 ms. In all 3 speech representations, i.e., word onsets, word surprisal

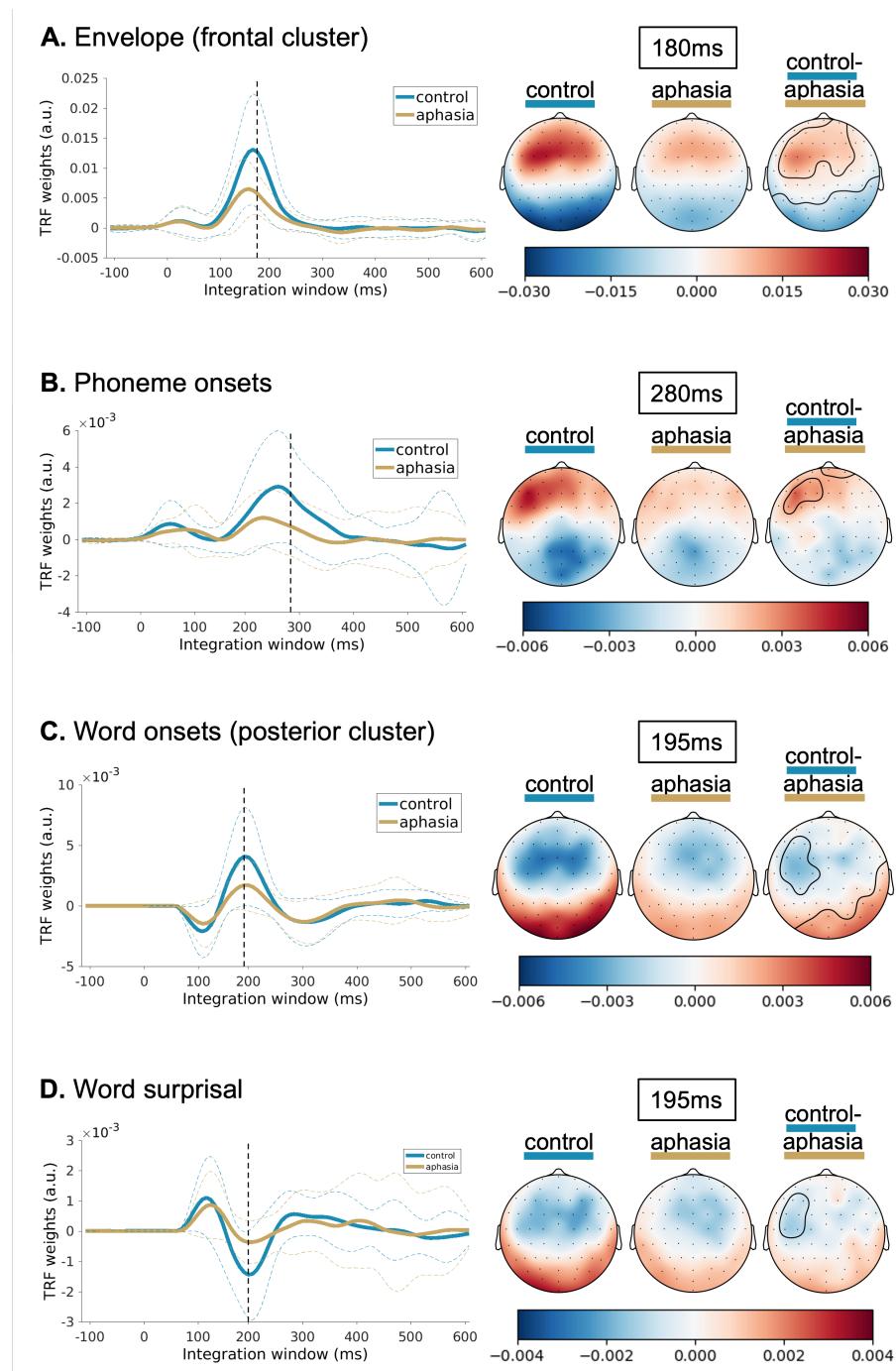


Figure 5: Individuals with aphasia show reduced neural response amplitudes to acoustic and linguistic speech representations in local clusters. In each of the 4 panels, the control and aphasia group average TRFs are plotted. For the left-sided figures, we averaged across the electrodes in the identified clusters. For the right-sided figures, the topoplots are shown at the time point indicated by the dashed line in the left-sided figures. The last topoplot on each panel displays the difference between the control group and the aphasia group, with the clusters being contoured. **A.** The neural response pattern to the speech envelope revealed 2 clusters around 180 ms, a frontal one (visualized) and a posterior one (not visualized). **B.** The neural response pattern to the phoneme onsets revealed a frontal cluster around 280 ms. **C.** The neural response pattern to the word onsets revealed 2 clusters around 195 ms, a frontal one (not visualized) and a posterior one (visualized). **D.** The neural response pattern to word surprisal revealed a cluster with 2 peaks, a posterior peak around 135 ms and a frontal peak around 195 ms (electrodes of this latter peak are visualized). Only the effect of the envelope was robust against correction for multiple comparisons.

509 and word frequency, these later clusters occurred over left-lateralized temporal electrodes (supplementary
510 fig. S.9). None of the other speech representations, i.e., phoneme onsets, phoneme surprisal and phoneme
511 entropy, showed any significantly different clusters, in neither of the 2 pair-wise group comparisons.

512 4 Discussion

513 In this study, we investigated whether IWA display different neural tracking of continuous speech than
514 age-matched healthy controls at multiple processing levels (acoustic to linguistic). To this end, we col-
515 lected EEG data of 39 IWA and 24 healthy controls while they listened to a continuous narrative. Speech
516 representations were derived from the narrative and their relation to the EEG signal studied. When the
517 neural signals align with the speech properties, this is referred to as neural tracking. This approach allows
518 to explore *how* spatio-temporal neural response patterns operate during continuous speech at multiple
519 processing levels (i.e., TRFs). Further, this method provides a measure of *how well* speech representa-
520 tions are encoded in the EEG data (i.e., prediction accuracy). Concerning prediction accuracies, group
521 differences between IWA and healthy controls were found for processing acoustic speech representations,
522 both located in posterior and frontal clusters (fig. 4). Regarding TRFs, group differences in processing
523 acoustic, segmentation and word-level speech representations were observed at different peaks between
524 135-280 ms, located in frontal, left-lateralized or posterior regions (fig. 5). However, as there were mul-
525 tiple TRF peaks identified per speech representation, we corrected for multiple comparisons and found
526 that only the clusters identified in the envelope TRF were robust group differences. Nonetheless, as this
527 is an exploratory study, we will discuss all findings here, because they can help to elicit more concrete
528 hypotheses in future investigations.

529 To date, neural tracking of acoustic and linguistic speech representations has not yet been investigated
530 in post-stroke aphasia. A study in individuals with PPA, a form of dementia that surfaces as language
531 impairment in its initial stages, has however reported increased speech envelope tracking in the theta
532 frequency band compared to healthy controls (Dial et al., 2021). The authors hypothesized that the
533 increased envelope tracking is related to the underlying physiological changes in PPA, i.e., a hypersyn-
534 chrony between frontal and temporo-parietal cortex as well as a hyperactivity in the frontal cortex (Dial
535 et al., 2021). Given the fundamentally different etiologies of aphasia after stroke and PPA, we did not
536 base any hypotheses on this study. Indeed, we found a rather contrasting pattern of results in the current
537 study, namely decreased acoustic neural tracking in IWA after stroke (fig. 4 A and B). We controlled
538 for the variance explained by segmentation and linguistic speech representations in the acoustic model
539 in order to yield a purer measure of acoustic processing. The observed group difference was also present
540 despite controlling for the variance explained by age, hearing, cognition, alertness and fatigue. Covariates
541 did not significantly explain any part of the variance in the acoustic model.

542 The envelope TRF confirms the finding of the acoustic prediction accuracy model, namely decreased
543 amplitudes in IWA over left frontal and posterior electrodes, additionally revealing that this difference
544 occurred around 180 ms neural processing time. These results may indicate that IWA track the slow
545 amplitude modulations of speech, which are essential for speech understanding (Shannon et al., 1995;

546 Zeng et al., 2005; Xu and Pfingst, 2008; Oganian and Chang, 2019), to a lesser extent than age-matched
547 healthy controls. Due to controlling the variance explained by the segmentation and linguistic speech
548 representations, the envelope TRF mainly contains the response to acoustic properties of speech. This
549 would mean that IWA also show impaired processing of speech acoustics. This is in line with findings
550 from Kries et al. (2023), who found that 76% of IWA had a low-level acoustic or phonemic processing
551 impairment.

552 At the sub-lexical and lexical segmentation level, we did not find a significant group difference. However,
553 the TRF analysis revealed locally decreased amplitudes in the neural response to phoneme onsets in
554 IWA at a peak around 280 ms as well as in the neural response to word onsets at a peak around 195
555 ms. These findings may indicate that IWA have a decreased performance when it comes to parsing
556 continuous speech. Humans segment continuous speech by making use of strategies such as analysis of
557 prosodic contours (bottom-up processes), making use of knowledge about distributional processing of
558 phonological information and about statistical regularities in word-to-object co-occurrences (top-down
559 processes) (Smith and Yu, 2008; Thiessen and Erickson, 2013; Suanda et al., 2014; Fló et al., 2019).
560 Thus, sub-lexical and lexical segmentation of continuous speech is a product of the interplay between
561 bottom-up and top-down processes (David et al., 2007; Shuai et al., 2014; Gaspers et al., 2017). Based
562 on the current findings, these processes that are necessary for speech segmentation may be impaired in
563 IWA.

564 EEG experiments that examined ERPs have found the N400 to be related to semantic activation and
565 integration into the sentence context (Kuperberg and Jaeger, 2016). N400 studies in IWA reported that
566 IWA in comparison to healthy controls have increased latencies of the N400 (Chang et al., 2016; Kawohl
567 et al., 2010; Khachatryan et al., 2017; Sheppard et al., 2017; Lice and Palmović, 2017) and attenuated
568 amplitudes (Robson et al., 2017; Sheppard et al., 2017; Kielar et al., 2012; Räling et al., 2016; Lice and
569 Palmović, 2017). Given the similarity between the N400 effects and the neural response pattern to word
570 surprisal (Michaelov and Bergen, 2020; Lopopolo and Rabovsky, 2022), we hypothesized to see differences
571 at the linguistic word level (i.e., word surprisal and word frequency) at a time window between 350 and
572 450 ms. However, we did not find a group difference in the time window of the N400 in the word surprisal
573 and word frequency TRFs. This may be due to the fact that older adults generally seem to have reduced
574 neural tracking of linguistic speech representations (Gillis and Kries et al., 2023). All participants in the
575 current study, healthy age-matched controls as well as IWA, were on average 70 years old. Older adults
576 may use different strategies to process lexical meaning than those captured by word surprisal and word
577 frequency (Federmeier et al., 2002; Wlotko et al., 2010; Spreng and Turner, 2019; Gillis and Kries et al.,
578 2023). More research is needed to determine an ideal set of speech representations to capture semantic
579 processing in healthy older adults, which can subsequently be translated to aphasia research. Another
580 option that may explain why no group difference was found in the target time window of word-level

581 contextual speech properties (i.e., 350-450 ms) is that the heterogeneity within the aphasia group may
582 have masked potential effects. Our exploratory, supplementary analysis with 2 aphasia subgroups - one
583 with milder or compensated language impairments and one with more severe language impairments -
584 displayed significant differences between the control group and the more severe aphasia group in word
585 surprisal and word frequency TRFs at 360-370 ms, thus falling within the N400 time window.

586 Surprisingly, we found that IWA have decreased amplitudes in the word surprisal TRF around 200 ms.
587 This peak is in line results from Gillis and Kries et al. (2023), where it was present in older, but not
588 younger adults. This peak may be related to lexical segmentation, since the word onsets TRF also showed
589 a group difference at the same latency and over similar electrodes. While we regressed out the influence
590 of word onsets to analyze the response to word surprisal and word frequency and vice-versa, the pulses
591 in the word surprisal and word frequency speech representations were set at the beginning of the word
592 and thus inherently also relate to a certain extent to lexical segmentation.

593 The IWA that participated in this study all had a left-hemispheric or bilateral lesion caused by stroke
594 and the data was collected in the chronic phase after stroke (i.e., ≥ 6 months), the median time that had
595 passed since stroke onset being 16 months. Stroke can cause changes in cerebral blood flow (Rabiller
596 et al., 2015; Brumm et al., 2010) and can create cavities filled with cerebro-spinal fluid at the lesion
597 site (Zbesko et al., 2018; Piastra et al., 2022). Especially the latter mechanism prevails also into the
598 chronic phase after stroke (Piastra et al., 2022; Salinet et al., 2014; Zbesko et al., 2018). Both aspects
599 can impact the conductivity of the neurophysiological activity that is picked up by the EEG, which can
600 lead to asymmetries in topography or changes in specific frequency bands (Vorwerk et al., 2014; Cohen
601 et al., 2015; Park et al., 2016; Cassidy et al., 2020). Therefore, in the following paragraph, we will discuss
602 whether the decreased neural tracking effects that we observed in IWA may be related to these underlying
603 anatomical changes that occur after stroke and influence the EEG signal or whether they are indeed a
604 trace of decreased processing of speech.

605 Many of the clusters that differed between healthy controls and IWA occurred over left-sided electrodes
606 (fig. 4 and 5). This region coincides with the region that is impacted by left-hemispheric lesions in the
607 area supplied by the middle cerebral artery, which is the lesion location for 79% of IWA in this study (fig.
608 1; supplementary table S.1). Thus, the lesion and hence, the altered conductivity, may influence the EEG
609 signal over these electrodes. While the EEG analysis method that we used here is not a direct measure of
610 the raw EEG signal - as is the case for ERPs - the altered conductivity may still affect measures of neural
611 tracking over lesion sites. The prediction accuracies are computed as the correlation between the recorded
612 EEG signal and the EEG signal predicted via modeling of the TRF and the speech representation. This
613 correlation does not take into account the absolute amplitude of the signals it compares. However, since
614 the signal-to-noise ratio of the EEG signal recorded over the lesion site is most likely lower than for other

615 electrodes, the prediction it makes will be more noisy too, resulting in a lower prediction accuracy. TRF
616 amplitudes are additionally more directly influenced by the recorded EEG amplitude. Thus, theoretically
617 neural tracking of speech properties can be impacted by the lesion-induced changes in conductivity of the
618 neurophysiological activity. We tested whether we would find a group difference when the lesion size is
619 regressed out for the aphasia group (section 3.2). We found that if IWA had no lesion, then they would
620 not differ from controls. We also found for the phoneme onsets model that the larger the lesion size is,
621 the lower the neural tracking scores within the aphasia group are. Together with the main results, this
622 indicates that the altered conductivity over the stroke lesion site, and in part probably also a difference
623 in processing slow amplitude modulations of speech may jointly explain the lower neural tracking in the
624 aphasia group. It should however be noted that this study was not designed to determine what factors
625 cause the group difference in neural tracking.

626 Some of the clusters in which the neural response amplitude significantly differs between groups (fig.
627 4B and 5) occurred over posterior electrodes. We can only speculate about these posterior clusters. As
628 shown in supplementary section S.1.5, IWA with a posterior lesion do not explain the occurrence of the
629 posterior clusters (supplementary fig. S.11). One possibility is that the large lesion overlap of IWA in
630 the left inferior frontal gyrus (fig. 1), mixed with lower speech processing-related activity in those areas,
631 led to lower measures of neural tracking over those electrodes. This resulted in a slight asymmetry with
632 higher amplitudes over anterior right-sided electrodes. The control group on the other hand showed a
633 slight asymmetry with higher amplitudes over anterior left-sided electrodes. These asymmetries in both
634 groups slightly change the dipole orientation in opposite directions. When the aphasia group topography
635 is subtracted from the control group topography, it consequently results in larger differences over anterior
636 left-sided electrodes and over posterior right-sided electrodes.

637 5 Conclusions and future outlook

638 In sum, our results show that measurements of neural tracking to specific speech properties may be a
639 promising avenue for future diagnostic and therapeutic applications. Especially the decrease in IWA in
640 processing of acoustic cues, such as the amplitude fluctuations and the onsets of phonemes and words,
641 seems to be a robust effect. Future studies may confirm the potential role of neural tracking of acoustic
642 and linguistic speech representations to provide profiles of language processing difficulties in IWA (i.e.,
643 acoustic, phonological, semantic). However, confounding factors, such as effects of the stroke lesion on the
644 EEG signal need to be taken into account in future investigations of neural tracking of speech in aphasia.
645 An interesting approach to address this issue could be the recruitment of a control group consisting of
646 individuals with a stroke without aphasia.

647 Once neural tracking in aphasia is better understood, the application potential of this method could be

648 manifold. A study that is currently in preparation found that IWA can be distinguished from healthy
649 controls with 83% accuracy based on neural envelope tracking with mutual information of only 5 to 7
650 minutes of EEG data (De Clercq et al., 2023). Looking a step further into the future, an aphasia diagnosis
651 based on processing levels of different speech representations could complement behavioral diagnosis and
652 inform speech-language pathologists further which functions should be trained in therapy. Test-retest
653 practice effects (i.e., improved performance on repeated tests due to remembering items or training test-
654 specific skills) could be avoided during therapy follow-up. Moreover, the method could be useful in the
655 acute phase after stroke, when behavioral diagnostic tests are too exhausting for patients. This would
656 still have to be tested in a clinically more compatible experimental setting in the future, but work by De
657 Clercq et al. (2023) shows that only few minutes of recording time would be needed to get reliable data.
658 Additionally, articulatory speech representations (e.g., mouth aperture, tongue protrusion (Mitra et al.,
659 2010)) could be investigated to analyze effects of production impairments in aphasia during listening.
660 Further, studying neural processing during speech production, using the same analytical framework, also
661 offers a possibility to study fluency impairments in IWA. Due to the feasibility of EEG in the clinical
662 context, the efficiency of the paradigm and the versatility of applications, examining neural tracking of
663 naturalistic, continuous speech provides a powerful approach to studying aphasia.

664 References

- 665 Aerts, A., van Mierlo, P., Hartsuiker, R. J., Santens, P., and De Letter, M. (2015). Neurophysiological
666 sensitivity for impaired phonological processing in the acute stage of aphasia. *Brain and Language*,
667 149:84–96.
- 668 Aiken, S. J. and Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear and Hearing*,
669 29(2):139–157.
- 670 Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. K. (2020). Gorilla in our
671 midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1):388–407.
- 672 Becker, F. and Reinvang, I. (2007). Successful syllable detection in aphasia despite processing impairments
673 as revealed by event-related potentials. *Behavioral and Brain Functions*, 3:1–16.
- 674 Biesmans, W., Das, N., Francart, T., and Bertrand, A. (2017). Auditory-inspired speech envelope extrac-
675 tion methods for improved eeg-based auditory attention detection in a cocktail party scenario. *IEEE*
676 *Transactions on Neural Systems and Rehabilitation Engineering*, 25(5):402–412.
- 677 Brady, M. C. (2022). Dosage, intensity, and frequency of language therapy for aphasia: A systematic
678 review-based, individual participant data network meta-analysis. *Stroke*, 29:956–967.
- 679 Brodbeck, C. (2020). Eelbrain 0.32. <http://doi.org/10.5281/zenodo.3923991>.
- 680 Brodbeck, C., Das, P., Gillis, M., Kulasingham, J. P., Bhattacharjee, S., Gaston, P., Resnik, P., and Simon,
681 J. Z. (2021). Eelbrain: A python toolkit for time-continuous analysis with temporal response functions.
682 *BioRxiv*.
- 683 Brodbeck, C., Presacco, A., and Simon, J. Z. (2018). Neural source dynamics of brain responses to
684 continuous stimuli: Speech processing from acoustics to comprehension. *NeuroImage*, 172:162–174.
- 685 Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., and Lalor, E. C. (2018). Electro-
686 physiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech.
687 *Current Biology*, 28:803–809.e3.
- 688 Broderick, M. P., Anderson, A. J., and Lalor, E. C. (2019). Semantic context enhances the early auditory
689 encoding of natural speech. *Journal of Neuroscience*, 39(38):7564–7575.
- 690 Broderick, M. P., Di Liberto, G. P., Anderson, A. J., Rofes, A., and Lalor, E. C. (2021). Dissociable elec-
691 trophysiological measures of natural language processing reveal differences in comprehension strategy
692 in healthy ageing. *Scientific reports*, 11(4963):1–12.

- 693 Brumm, K. P., Perthen, J. E., Liu, T. T., Haist, F., Ayalon, L., and Love, T. (2010). An arterial
694 spin labeling investigation of cerebral blood flow deficits in chronic stroke survivors. *NeuroImage*,
695 51(2010):995–1005.
- 696 Cassidy, J. M., Wodeyar, A., Wu, J., Kaur, K., Masuda, A. K., Srinivasan, R., and Cramer, S. C. (2020).
697 Low-frequency oscillations are a biomarker of injury and recovery after stroke. *Stroke*, 51(5):1442–1450.
- 698 Chang, C. T., Lee, C. Y., Chou, C. J., Fuh, J. L., and Wu, H. C. (2016). Predictability effect on N400
699 reflects the severity of reading comprehension deficits in aphasia. *Neuropsychologia*, 81:117–128.
- 700 Cocquyt, E.-M., Vandewiele, M., Bonnarens, C., Santens, P., and De Letter, M. (2020). The sensitivity of
701 event-related potentials/fields to logopedic interventions in patients with stroke-related aphasia. *Acta*
702 *Neurologica Belgica*, 120(4):2240–2993.
- 703 Cohen, R., Abboud, S., and Arad, M. (2015). Monitoring brain damage using bioimpedance technique
704 in a 3d numerical model of the head. *Medical Engineering and Physics*, 37(5):453–459.
- 705 Cordella, C., Munsell, M., Godlove, J., Anantha, V., Advani, M., and Kiran, S. (2022). Dosage frequency
706 effects on treatment outcomes following self-managed digital therapy: Retrospective cohort study.
707 *Journal of Medical Internet Research*, 24(7):1–14.
- 708 David, S. V., Mesgarani, N., and Shamma, S. A. (2007). Estimating sparse spectro-temporal receptive
709 fields with natural stimuli. *Network: Computation in Neural Systems*, 18:191–212.
- 710 De Clercq, P., Kries, J., Mehraram, R., Vanthornhout, J., Francart, T., and Vandermosten, M. (2023).
711 Detecting post-stroke aphasia using eeg-based neural envelope tracking of natural speech. *medrxiv*.
- 712 Di Liberto, G., Peter, V., Kalashnikova, M., Goswami, U., Burnham, D., and Lalor, E. (2018). Atyp-
713 ical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia.
714 *NeuroImage*, 175:70–79.
- 715 Di Liberto, G. M. and Lalor, E. C. (2017). Indexing cortical entrainment to natural speech at the
716 phonemic level : Methodological considerations for applied research. *Hearing Research*, 348:70–77.
- 717 Di Liberto, G. M., O’Sullivan, J. A., and Lalor, E. C. (2015). Low frequency cortical entrainment to
718 speech reflects phoneme level processing. *Current Biology*, 25:2457–2465.
- 719 Dial, H. R., Gnanateja, G. N., Tessmer, R. S., Gorno-Tempini, M. L., Chandrasekaran, B., and Henry,
720 M. L. (2021). Cortical tracking of the speech envelope in logopenic variant primary progressive aphasia.
721 *Frontiers in Human Neuroscience*, 14(597694):1–9.
- 722 Ding, N. and Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural
723 and dichotic listening. *Journal of Neurophysiology*, 107:78–89.

- 724 Duchateau, J., Kong, Y. O., Cleuren, L., Latacz, L., Roelens, J., Samir, A., Demuynck, K., Ghesquière,
725 P., Verhelst, W., and Van hamme, H. (2009). Developing a reading tutor: Design and evaluation of
726 dedicated speech recognition and synthesis modules. *Speech Communication*, 51(10):985–994.
- 727 El Hachioui, H., Visch-Brink, E. G., Lingsma, H. F., Van De Sandt-Koenderman, M. W., Dippel, D. W.,
728 Koudstaal, P. J., and Middelkoop, H. A. (2014). Nonlinguistic cognitive impairment in poststroke
729 aphasia: A prospective study. *Neurorehabilitation and Neural Repair*, 28(3):273–281.
- 730 Engelter, S. T., Gostynski, M., Papa, S., Frei, M., Born, C., Ajdacic-Gross, V., Gutzwiller, F., and Lyrer,
731 P. A. (2006). Epidemiology of aphasia attributable to first ischemic stroke: Incidence, severity, fluency,
732 etiology, and thrombolysis. *Stroke*, 37(6):1379–1384.
- 733 Federmeier, K. D., McLennan, D. B., de Ochoa, E., and Kutas, M. (2002). The impact of semantic
734 memory organization and sentence context information on spoken language processing by younger and
735 older adults: An ERP study. *Psychophysiology*, 39(2):133–146.
- 736 Flamand-Roze, C., Falissard, B., Roze, E., Maintigneux, L., Beziz, J., Chacon, A., Join-Lambert, C.,
737 Adams, D., and Denier, C. (2011). Validation of a new language screening tool for patients with acute
738 stroke. *Stroke*, 42(5):1224–1229.
- 739 Fló, A., Brusini, P., Macagno, F., Nespor, M., Mehler, J., and Ferry, A. L. (2019). Newborns are sensitive
740 to multiple cues for word segmentation in continuous speech. *Developmental Science*, 22(4):1–16.
- 741 Fonseca, J., Raposo, A., and Martins, I. P. (2018). Cognitive functioning in chronic post-stroke aphasia.
742 *Applied Neuropsychology: Adult*, 26(4):355–364.
- 743 Francart, T., Wieringen, A. V., and Wouters, J. (2008). Apex 3: a multi-purpose test platform for
744 auditory psychophysical experiments. *J Neurosci Methods*, 172:283–293.
- 745 Fridriksson, J., Ouden, D. B. D., Hillis, A. E., Hickok, G., Rorden, C., Basilakos, A., Yourganov, G., and
746 Bonilha, L. (2018). Anatomy of aphasia revisited. *Brain*, 141:848–862.
- 747 Gaspers, J., Cimiano, P., Rohlfing, K., and Wrede, B. (2017). Constructing a language from scratch:
748 Combining bottom-up and top-down learning processes in a computational model of language acqui-
749 sition. *IEEE Transactions on Cognitive and Developmental Systems*, 9(2):183–196.
- 750 Gillis, M., Van Canneyt, J., Francart, T., and Vanthornhout, J. (2022). Neural tracking as a diagnostic
751 tool to assess the auditory pathway. *Hearing Research*, 426(108607):1–14.
- 752 Gillis, M., Vanthornhout, J., Simon, J. Z., Francart, T., and Brodbeck, C. (2021). Neural markers
753 of speech comprehension: measuring eeg tracking of linguistic speech representations, controlling the
754 speech acoustics. *Journal of Neuroscience*, 41(50):10316–10329.

- 755 Gillis and Kries, Vandermosten, M., and Francart, T. (2023). Neural tracking of linguistic and acoustic
756 speech representations decreases with advancing age. *NeuroImage*, 267(119841):1–16.
- 757 Hamilton, L. S. and Huth, A. G. (2018). The revolution will not be controlled: natural stimuli in speech
758 neuroscience. *Language, Cognition and Neuroscience*, 35(5):573–582.
- 759 Harris, K. C. (2020). The Aging Auditory System: Electrophysiology. In Helfer, K. S., Bartlett, E. L.,
760 Popper, A. N., and Fay, R. R., editors, *Aging and Hearing: Causes and Consequences*, pages 117–141.
761 Springer International Publishing, Cham.
- 762 Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews
763 Neuroscience*, 8(May):393–402.
- 764 Hillyard, S. and Kutas, M. (1984). Brain potentials during reading reflect word expectancy and semantic
765 association. *Nature*, 307(5947):161–163.
- 766 Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., and Theunissen, F. E. (2017). En-
767 coding and decoding models incognitive electrophysiology. *Frontiers in Systems Neuroscience*, 11(61):1–
768 24.
- 769 Huygelier, H., Schraepen, B., Demeyere, N., and Gillebert, C. R. (2019). The Dutch version of the Oxford
770 Cognitive Screen (OCS-NL): normative data and their association with age and socio-economic status.
771 *Aging, Neuropsychology, and Cognition*, 27(5):765–786.
- 772 Ilvonen, T., Kujala, T., Kozou, H., Kiesiläinen, A., Salonen, O., Alku, P., and Näätänen, R. (2004). The
773 processing of speech and non-speech sounds in aphasic patients as reflected by the mismatch negativity
774 (MMN). *Neuroscience Letters*, 366(3):235–240.
- 775 Ilvonen, T.-M., Kujala, T., Tervaniemi, M., Salonen, O., Näätänen, R., and Pekkonen, E. (2001). The
776 processing of sound duration after left hemisphere stroke: Event-related potential and behavioral evi-
777 dence. *Psychophysiology*, 38:622–628.
- 778 Kandylaki, K. D. and Bornkessel-Schlesewsky, I. (2019). From story comprehension to the neurobiology
779 of language. *Language, Cognition and Neuroscience*, 34(4):405–410.
- 780 Kawohl, W., Bunse, S., Willmes, K., Hoffrogge, A., Buchner, H., and Huber, W. (2010). Semantic event-
781 related potential components reflect severity of comprehension deficits in aphasia. *Neurorehabilitation
782 and Neural Repair*, 24(3):282–289.
- 783 Keuleers, E., Brysbaert, M., and New, B. (2010). Sublex-nl: A new measure for dutch word frequency
784 based on film subtitles. *Behavior research methods*, 42(3):643–650.

- 785 Khachatryan, E., De Letter, M., Vanhoof, G., Goeleven, A., and Van Hulle, M. M. (2017). Sentence
786 context prevails over word association in aphasia patients with spared comprehension: Evidence from
787 N400 event-related potential. *Frontiers in Human Neuroscience*, 10(684):1–15.
- 788 Kielar, A., Meltzer-Asscher, A., and Thompson, C. K. (2012). Electrophysiological responses to argument
789 structure violations in healthy adults and individuals with agrammatic aphasia. *Neuropsychologia*,
790 50(14):3320–3337.
- 791 Kries, J., Clercq, P. D., Lemmens, R., Francart, T., and Vandermosten, M. (2023). Acoustic and phonemic
792 processing are impaired in individuals with aphasia. *Scientific Reports*, 13.
- 793 Kuperberg, G. R. and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?
794 *Language, Cognition and Neuroscience*, 31(1):32–59.
- 795 Kutas, M. and Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400
796 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, 62(1):621–
797 647.
- 798 Lalor, E. C. and Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted
799 with precise temporal resolution. *European Journal of Neuroscience*, 31:189–193.
- 800 Le, D., Licata, K., and Mower Provost, E. (2018). Automatic quantitative analysis of spontaneous aphasic
801 speech. *Speech Communication*, 100:1–12.
- 802 Lesenfants, D. and Francart, T. (2020). The interplay of top-down focal attention and the cortical tracking
803 of speech. *Scientific Reports*, 10(6922):1–10.
- 804 Lice, K. and Palmović, M. (2017). Semantic categorization in aphasic patients with impaired language
805 comprehension: An event-related potentials study. *Suvremena lingvistika*, 43(84):135–155.
- 806 Lopopolo, A. and Rabovsky, M. (2022). Tracking lexical and semantic prediction error underlying the
807 n400 using artificial neural network models of sentence processing. *biorxiv*.
- 808 Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., and Zion Golumbic, E. (2017). Sleep
809 disrupts high-level speech parsing despite significant basic auditory processing. *Journal of Neuroscience*,
810 37:7772–7781.
- 811 Martin, B. A., Tremblay, K. L., and Korczak, P. (2008). Speech Evoked Potentials : From the Laboratory
812 to the Clinic. *Ear and Hearing*, 29:285–313.
- 813 MATLAB (2016). *version 9.1.0.441655 (R2016b)*. The MathWorks Inc., Natick, Massachusetts.
- 814 Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human
815 superior temporal gyrus. *Science*, 343(2):1006–1010.

- 816 Mesik, J., Ray, L., and Wojtczak, M. (2021). Effects of Age on Cortical Tracking of Word-Level Features
817 of Continuous Competing Speech. *Frontiers in Neuroscience*, 15(4):1–21.
- 818 Michaelov, J. A. and Bergen, B. K. (2020). How well does surprisal explain n400 amplitude under different
819 experimental conditions? In *Proceedings of the 24th Conference on Computational Natural Language
820 Learning*, pages 652–663, online. Association for Computational Linguistics.
- 821 Mitra, V., Nam, H., Espy-Wilson, C. Y., Saltzman, E., and Goldstein, L. (2010). Retrieving tract
822 variables from acoustics: A comparison of different machine learning strategies. *IEEE J Sel Top Signal
823 Process*, 4:1027–1045.
- 824 National Aphasia Association (Accessed in August 2022). Aphasia fact sheet. *NAA online publication*.
- 825 Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson,
826 H. J., Fu, X., Heyselaar, E., Huettig, F., Husband, E. M., Ito, A., Kazanina, N., Kogan, V., Kohút,
827 Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., Rueschemeyer, S. A., Segaert, K.,
828 Tuomainen, J., and Von Grebmer Zu Wolfsturn, S. (2020). Dissociable effects of prediction and
829 integration during language comprehension: Evidence from a largescale study using brain potentials.
830 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1791):1–9.
- 831 Ofek, E., Purdy, S. C., Ali, G., Webster, T., Gharahdaghi, N., and McCann, C. M. (2013). Processing of
832 emotional words after stroke: An electrophysiological study. *Clinical Neurophysiology*, 124:1771–1778.
- 833 Oganian, Y. and Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human
834 superior temporal gyrus. *Science Advances*, 5(11):1–13.
- 835 Oldfield, R. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsy-
836 chologia*, 9(1):97–113.
- 837 Park, W., Kwon, G. H., Kim, Y. H., Lee, J. H., and Kim, L. (2016). Eeg response varies with lesion
838 location in patients with chronic stroke. *Journal of NeuroEngineering and Rehabilitation*, 13(21):1–10.
- 839 Pasley, B. N. and Knight, R. T. (2013). Decoding speech for understanding and treating aphasia. *Progress
840 in Brain Research*, 207:435–456.
- 841 Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory
842 cortex are enhanced during comprehension. *Cerebral Cortex*, 23(6):1378–1387.
- 843 Pettigrew, C. M., Murdoch, B. E., Kei, J., Ponton, C. W., Alku, P., and Chenery, H. J. (2005). The
844 mismatch negativity (MMN) response to complex tones' and spoken words in individuals with aphasia.
845 *Aphasiology*, 19(2):131–163.

- 846 Piastra, M. C., Oostenveld, R., Schoffelen, J. M., and Piai, V. (2022). Estimating the influence of stroke
847 lesions on meg source reconstruction. *NeuroImage*, 260:1–13.
- 848 Prinsloo, K. D. and Lalor, E. C. (2022). General auditory and speech-specific contributions to cortical
849 envelope tracking revealed using auditory chimeras. *Journal of Neuroscience*, 42(41):7782–7798.
- 850 Pulvermüller, F., Mohr, B., and Lutzenberger, W. (2004). Neurophysiological correlates of word and
851 pseudo-word processing in well-recovered aphasics and patients with right-hemispheric stroke. *Psy-
852 chophysiology*, 41(4):584–591.
- 853 R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for
854 Statistical Computing, Vienna, Austria.
- 855 Rabiller, G., He, J.-W., Nishijima, Y., Wong, A., and Liu, J. (2015). Perturbation of brain oscillations
856 after ischemic stroke: A potential biomarker for post-stroke function and therapy. *Int. J. Mol. Sci.*,
857 16:25605–25640.
- 858 Räling, R., Schröder, A., and Wartenburger, I. (2016). The origins of age of acquisition and typicality
859 effects: Semantic processing in aphasia and the ageing brain. *Neuropsychologia*, 86:80–92.
- 860 Robson, H., Pilkington, E., Evans, L., DeLuca, V., and Keidel, J. L. (2017). Phonological and semantic
861 processing during comprehension in Wernicke's aphasia: An N400 and Phonological Mapping Negativ-
862 ity Study. *Neuropsychologia*, 100:144–154.
- 863 Rohde, A., Worrall, L., Godecke, E., O'Halloran, R., Farrell, A., and Massey, M. (2018). Diagnosis of
864 aphasia in stroke populations: A systematic review of language tests. *PLoS ONE*, 13(3):1–17.
- 865 Salinet, A. S. M., Panerai, R. B., Robinson, T. G., and Salinet, A. (2014). The longitudinal evolution of
866 cerebral blood flow regulation after acute ischaemic stroke. *Cerebrovasc Dis Extra*, 4:186–197.
- 867 Schevenels, K., Gerrits, R., Lemmens, R., De Smedt, B., Zink, I., and Vandermosten, M. (2022). Early
868 white matter connectivity and plasticity in post stroke aphasia recovery. *NeuroImage: Clinical*, 36:1–12.
- 869 Schevenels, K., Price, C. J., Zink, I., De Smedt, B., and Vandermosten, M. (2020). A review on treatment-
870 related brain changes in aphasia. *Neurobiology of Language*, 1(4):402–433.
- 871 Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with
872 primarily temporal cues. *Science*, 270(5234):303–304.
- 873 Sheppard, S. M., Love, T., Midgley, K. J., Holcomb, P. J., and Shapiro, L. P. (2017). Electrophysiology of
874 prosodic and lexical-semantic processing during sentence comprehension in aphasia. *Neuropsychologia*,
875 107:9–24.

- 876 Shuai, L., Gong, T., Giuliano, R., and Gu, W. (2014). Temporal relation between top-down and bottom-
877 up processing in lexical tone perception. *Frontiers in Behavioral Neuroscience*, 8(97):1–16.
- 878 Silkes, J. A. P. and Anjum, J. (2021). The role and use of event-related potentials in aphasia: A scoping
879 review. *Brain and Language*, 219:1–16.
- 880 Smith, L. and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics.
881 *Cognition*, 106(3):1558–1568.
- 882 Somers, B., Francart, T., and Bertrand, A. (2018). A generic eeg artifact removal algorithm based on
883 the multi-channel wiener filter. *Journal of Neural Engineering*, 15(3):1–13.
- 884 Spreng, R. N. and Turner, G. R. (2019). The Shifting Architecture of Cognition and Brain Function in
885 Older Adulthood. *Perspectives on Psychological Science*, 14(4):523–542.
- 886 Suanda, S. H., Mugwanya, N., and Namy, L. L. (2014). Cross-situational statistical word learning in
887 young children. *Journal of Experimental Child Psychology*, 126:395–411.
- 888 Thiessen, E. and Erickson, L. (2013). Discovering words in fluent speech: The contribution of two kinds
889 of statistical information. *Frontiers in Psychology*, 3:1–10.
- 890 Tremblay, P. and Dick, A. S. (2016). Broca and wernicke are dead, or moving past the classic model of
891 language neurobiology. *Brain and Language*, 162:60–71.
- 892 Van Ewijk, E., Dijkhuis, L., Hofs-Van Kats, M., Hendrickx-Jessurun, M., Wijngaarden, M., and De
893 Hilster, C. (2020). *Nederlandse Benoem Test*. Bohn Stafleu Van Loghum, Houten, NL.
- 894 Van Rossum, G. and Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en
895 Informatica Amsterdam.
- 896 Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., and Francart, T. (2018). Speech intelligibility
897 predicted from neural entrainment of the speech envelope. *JARO - Journal of the Association for
898 Research in Otolaryngology*, 19(2):181–191.
- 899 Verwimp, L., Van hamme, H., and Wambacq, P. (2019). Tf-lm: Tensorflow-based language modeling
900 toolkit. In <http://www.lrec-conf.org/proceedings/lrec2018/index.html>, pages 2968–2973. Proceedings
901 LREC.
- 902 Visch-Brink, E., Van de Sandt-Koenderman, M., and El Hachioui, H. (2010). *ScreeLing*. Houten: Bohn
903 Stafleu Van Loghum.
- 904 Vorwerk, J., Cho, J. H., Rampp, S., Hamer, H., Knösche, T. R., and Wolters, C. H. (2014). A guideline
905 for head volume conductor modeling in eeg and meg. *NeuroImage*, 100:590–607.

- 906 Weissbart, H., Kandylaki, K. D., and Reichenbach, T. (2019). Cortical tracking of surprisal during
907 continuous speech comprehension. *Journal of Cognitive Neuroscience*, 32(1):155–166.

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

911 Wlotko, E. W., Lee, C. L., and Federmeier, K. D. (2010). Language of the Aging Brain: Event-Related
912 Potential Studies of Comprehension in Older Adults. *Linguistics and Language Compass*, 4(8):623–638.

913 Xu, L. and Pfingst, B. E. (2008). Spectral and temporal cues for speech recognition: Implications for
914 auditory prostheses. *Hearing Research*, 242(1-2):132–140.

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

918 Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K.
919 (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National
920 Academy of Sciences of the United States of America*, 102(7):2293–2298.

921 List of Figures

| | | |
|-----|--|----|
| 934 | 4 Individuals with aphasia display decreased neural tracking of acoustic speech | |
| 935 | representations - across all electrodes and in local clusters. A. When averaging | |
| 936 | the prediction accuracies of the acoustic model across all 64 electrodes, we found a signif- | |
| 937 | icant group difference, even when we controlled for age, hearing, cognition, alertness and | |
| 938 | fatigue. The acoustic model consists of the speech envelope and its onsets as speech rep- | |
| 939 | resentations. B. The cluster-based permutation test revealed 2 clusters that significantly | |
| 940 | differed between groups, showing lower prediction accuracies in individuals with aphasia. | |
| 941 | The lowest topoplot consists of the difference between the control group and the aphasia | |
| 942 | group and the significant clusters are contoured. | 20 |
| 943 | 5 Individuals with aphasia show reduced neural response amplitudes to acoustic | |
| 944 | and linguistic speech representations in local clusters. In each of the 4 panels, the | |
| 945 | control and aphasia group average TRFs are plotted. For the left-sided figures, we averaged | |
| 946 | across the electrodes in the identified clusters. For the right-sided figures, the topoplots | |
| 947 | are shown at the time point indicated by the dashed line in the left-sided figures. The last | |
| 948 | topoplot on each panel displays the difference between the control group and the aphasia | |
| 949 | group, with the clusters being contoured. A. The neural response pattern to the speech | |
| 950 | envelope revealed 2 clusters around 180 ms, a frontal one (visualized) and a posterior one | |
| 951 | (not visualized). B. The neural response pattern to the phoneme onsets revealed a frontal | |
| 952 | cluster around 280 ms. C. The neural response pattern to the word onsets revealed 2 | |
| 953 | clusters around 195 ms, a frontal one (not visualized) and a posterior one (visualized). D. | |
| 954 | The neural response pattern to word surprisal revealed a cluster with 2 peaks, a posterior | |
| 955 | peak around 135 ms and a frontal peak around 195 ms (electrodes of this latter peak are | |
| 956 | visualized). Only the effect of the envelope was robust against correction for multiple | |
| 957 | comparisons. | 22 |
| 958 | S.1 Flowchart of procedure to recruit participants with aphasia. | 41 |
| 959 | S.2 Timeline of data collection of healthy controls and participants with aphasia. . . | 41 |
| 960 | S.3 Visualization of scores on language tests. The cut-off scores are marked by the | |
| 961 | dashed lines. | 43 |

| | | |
|-----|--|----|
| 962 | S.4 Subjective ratings of alertness and fatigue throughout the experimental session. | |
| 963 | A. and B. The alertness and fatigue questions as presented on a tablet to the participants, | |
| 964 | with visual icons to aid comprehension of the task. The questions were simultaneously | |
| 965 | presented auditorily. C. A linear mixed effects model showed a main effect of group, but | |
| 966 | no main effect of time point or interaction effect for the alertness scale. D. A linear mixed | |
| 967 | effects model showed a significant interaction effect between group and time point for the | |
| 968 | fatigue scale. E. Visualization of the experimental protocol with regard to the time points | |
| 969 | at which the alertness and fatigue scales were administered. | 45 |
| 970 | S.5 Time ranges of 150 ms visualized in 3 speech representations. | 46 |
| 971 | S.6 Exploratory correlation matrix. Only significant correlations are colored. Note: Rise | |
| 972 | time and phoneme identification tasks are measuring acoustic and phoneme processing, | |
| 973 | the tasks are described in detail in (Kries et al., 2023). MCQ=multiple choice question; | |
| 974 | YNQ=Yes/No question. | 47 |
| 975 | S.7 Visualization of scores on language tests used for splitting up the aphasia group | |
| 976 | into a milder and a more severe group. | 48 |
| 977 | S.8 Only the subgroup of aphasia with more severe language difficulties shows de- | |
| 978 | creased neural tracking of acoustic speech representations - across all electrodes | |
| 979 | and in local clusters. A. When averaging the prediction accuracies of the acoustic model | |
| 980 | across all 64 electrodes, we found a significant group effect, even when we controlled for | |
| 981 | age, hearing, cognition, alertness, fatigue and lesion size. The acoustic model consists of | |
| 982 | the speech envelope and its onsets as speech representations. B. The cluster-based per- | |
| 983 | mutation test for the group comparison between control and aphasia subgroup, but not | |
| 984 | between the control and mild aphasia subgroup, revealed 2 clusters that significantly dif- | |
| 985 | fered between groups, showing lower prediction accuracies in individuals with more severe | |
| 986 | aphasia. The lowest topoplot consists of the difference between the control group and the | |
| 987 | aphasia/mild aphasia group and the contours encircle the significant electrode clusters. . . | 49 |

| | | |
|------|---|----|
| 988 | S.9 Individuals with more severe aphasia show reduced neural response patterns to | |
| 989 | acoustic and linguistic representations in local clusters compared to controls. | |
| 990 | The top most topoplot of each panel shows the control group average and the middle plot | |
| 991 | shows the aphasia subgroup average. The spaghetti plots show the TRF weights across | |
| 992 | the integration window. Each line represents one of the 64 electrodes. The topoplots | |
| 993 | display the topography of the neural response at the time point indicated in the TRF plot | |
| 994 | with a vertical black line. The bottom most TRF plot and topoplot display the difference | |
| 995 | between the control group and the aphasia subgroup with more severe language problems. | |
| 996 | In the bottom TRF plot of each panel, the colored parts represent the electrodes within | |
| 997 | the significant cluster and the integration window over which they are significant. | 51 |
| 998 | S.10 Lesion volume of the mild aphasia and more severe aphasia group. The lesion | |
| 999 | volume is significantly larger in the more severe aphasia group as compared to the mild | |
| 1000 | aphasia group. | 52 |
| 1001 | S.11 The patterns resulting from the cluster-based permutations when the 3 IWA | |
| 1002 | with lesions resulting from PCA damage are excluded are similar to the pattern | |
| 1003 | of results when the whole aphasia group is compared to the control group. In | |
| 1004 | the encircled clusters, the aphasia group had significantly lower amplitudes than the control | |
| 1005 | group. | 53 |

1006 Supplementary material

1007 S.1.1 Recruitment, data collection and participant details

1008 S.1.1.1 Flowchart of recruitment procedure

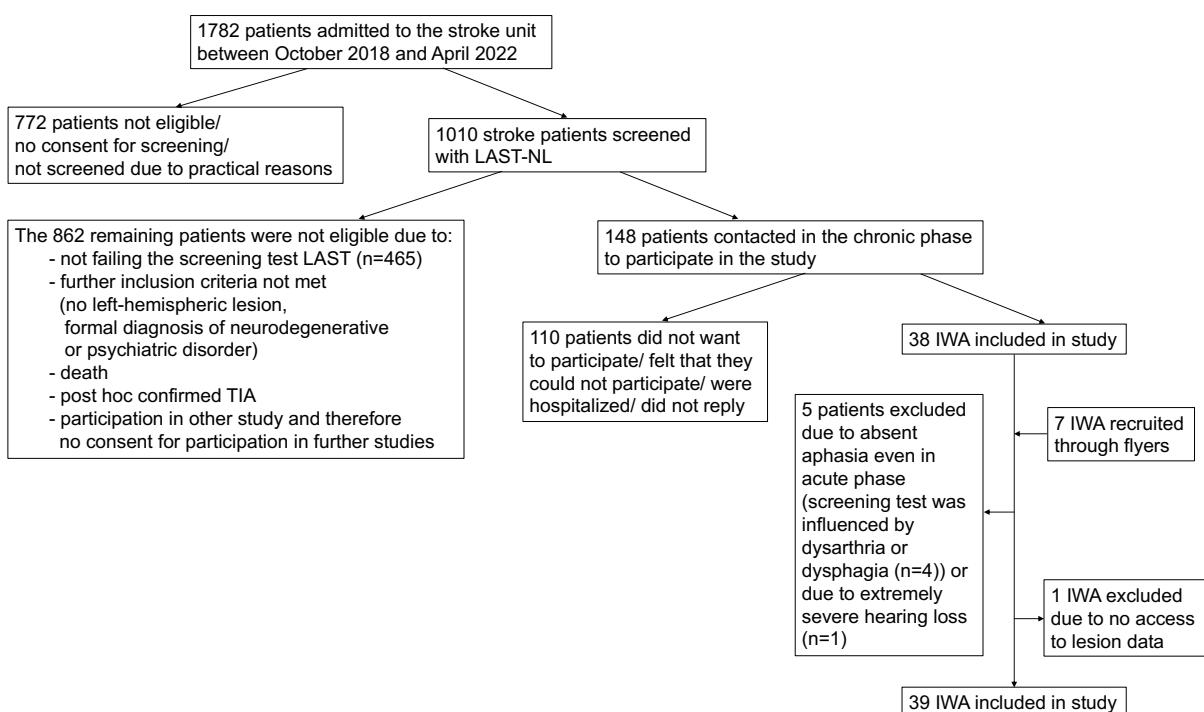


Figure S.1: Flowchart of procedure to recruit participants with aphasia.

1009 S.1.1.2 Timeline of data collection

1010 Data collection had to be stopped during winters of 2020/2021 and 2021/2022 due to the COVID-
1011 19 pandemic and thus took place mainly between the months June and October of each year. Healthy
1012 controls and participants with aphasia were recruited in alternation throughout the 3 years, which allowed
1013 us to age-match healthy controls to the IWA. Moreover this ensured that changes in EEG hardware did
1014 not influence one group more than the other. The electrode set was changed in September 2021 to prevent
future problems due to extensive usage of the previous set.

Data collection

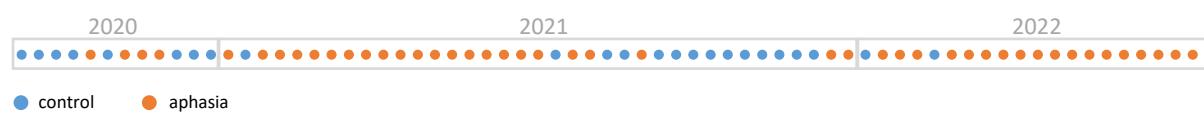


Figure S.2: Timeline of data collection of healthy controls and participants with aphasia.

1015

1016 S.1.1.3 Participant details

Table S.1: Demographic, lesion and diagnostic information of the aphasia group

| ID(n=39) | age | sex | Time since stroke (months) | Stroke type | Blood vessel | Lesioned hemisphere | 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

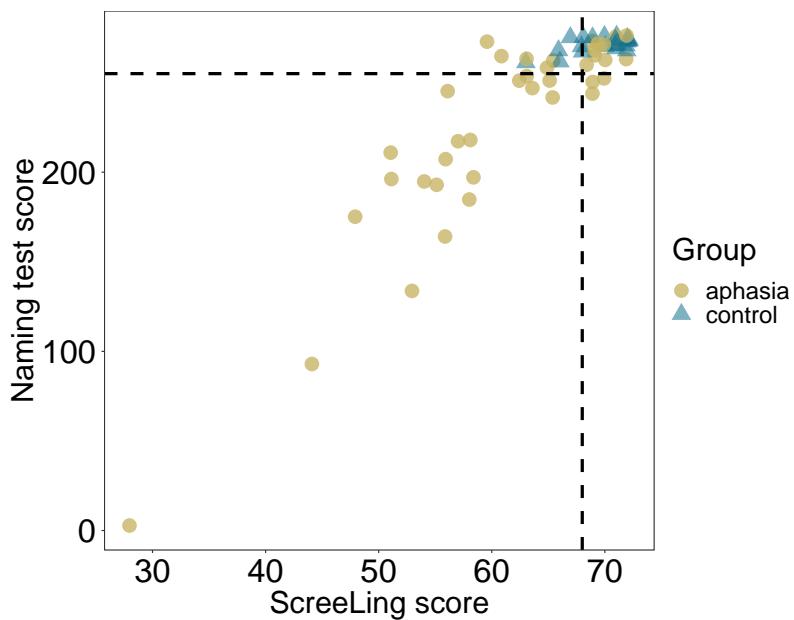


Figure S.3: Visualization of scores on language tests. The cut-off scores are marked by the dashed lines.

1017 **S.1.1.4 Lesion delineation**

1018 MRI data collection was not a part of the planned data collection for this study, but ethical agreement
1019 was given to access the scans from the medical files of the university hospital. Thus, we had access to
1020 T2-weighted FLAIR images that were administered in the acute stage post-stroke, based on which the
1021 segmentations of the affected stroke tissue were manually performed. For 10 participants, scans were
1022 available from the chronic phase post-stroke, as they took part in an MRI study (unpublished). 12
1023 participants had participated in an earlier aphasia study, and thus we could use the already established
1024 segmentation maps.

1025 **S.1.1.5 Monitoring alertness and fatigue throughout the experimental session**

1026 The subjective ratings of alertness and fatigue were administered in order to use them as covariates in
1027 statistical models. The EEG measurement took place between t1 and t2, therefore we averaged across
1028 t1 and t2 of the alertness and fatigue rating respectively, which were then used as covariate factors for
1029 the analysis of prediction accuracy models (fig. 4). We also analyzed the course of alertness and fatigue
1030 ratings between groups. Linear mixed effects models were employed for this purpose. The analysis of the
1031 alertness question revealed a main effect of group ($p < 0.001$), but no main effect of time point or interaction
1032 effect (time point: $p = 0.06$; interaction: $p = 0.226$; fig. S.4 C). Thus, across time points, individuals with
1033 aphasia were less alert than healthy controls. The analysis of the fatigue question showed significant
1034 main effects (group: $p < 0.001$; time point: $p = 0.015$) and a significant interaction effect between group

1035 and time point ($p=0.02$; fig. S.4 D). Post hoc testing revealed that while the initial level of fatigue was at
1036 a similar level for both groups at the beginning of the test session (t_1 ; $p=0.92$), individuals with aphasia
1037 were more tired than healthy controls after the first part of the experimental session (t_2 ; $p<0.001$) and
1038 at the end of the protocol (t_3 ; $p=0.002$).

1039 We offered IWA to continue the administration of the second part of the experimental protocol on another
1040 day in case they felt too exhausted after the EEG measurement (fig. S.4 E). Indeed, 5 out of 41 IWA felt
1041 too exhausted at t_2 and chose to do the rest of the tests on another day. This second session took place
1042 at the patients' homes, so that they did not have to arrange transportation to the lab again.

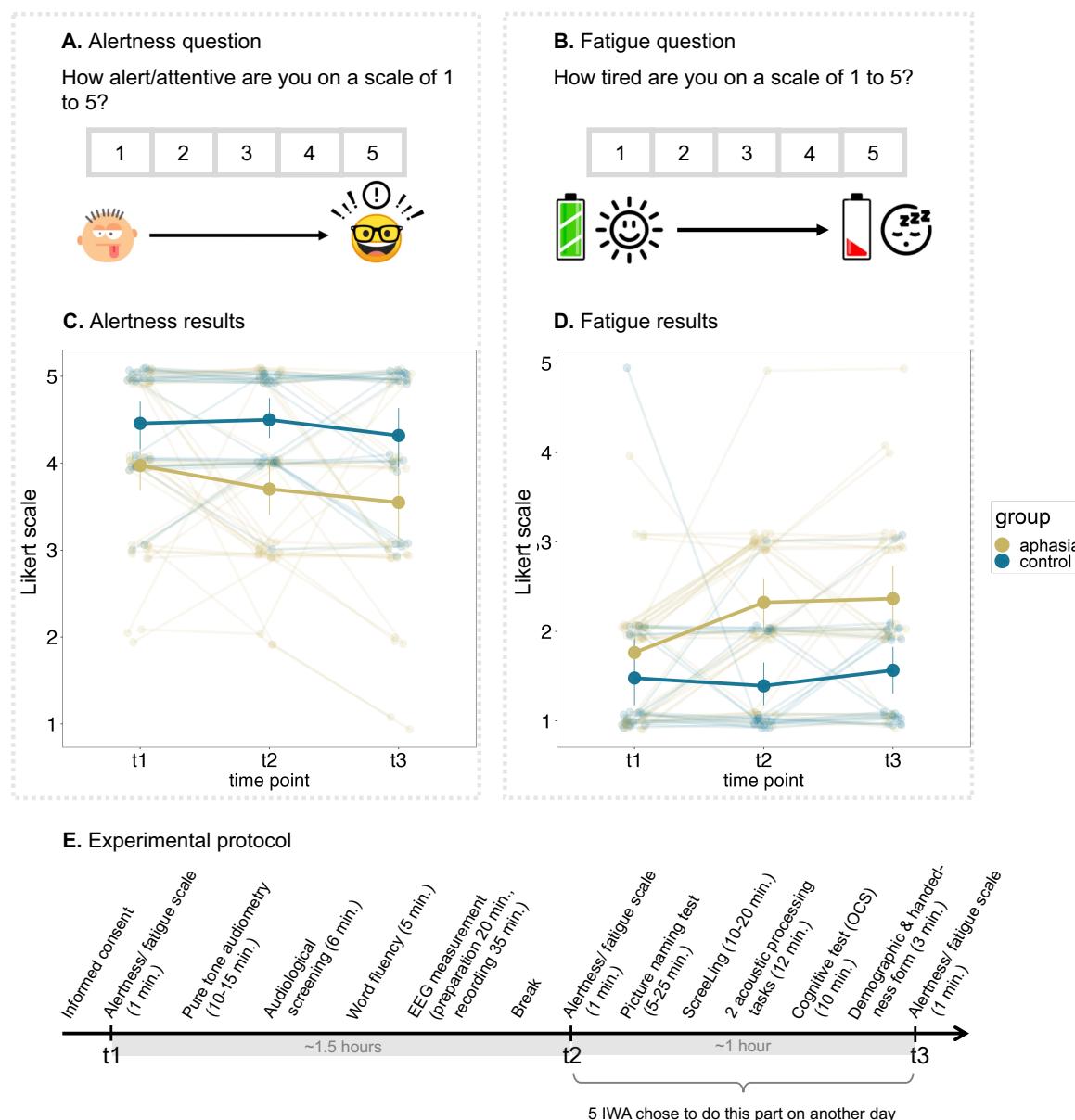


Figure S.4: Subjective ratings of alertness and fatigue throughout the experimental session.
A. and B. The alertness and fatigue questions as presented on a tablet to the participants, with visual icons to aid comprehension of the task. The questions were simultaneously presented auditorily. **C.** A linear mixed effects model showed a main effect of group, but no main effect of time point or interaction effect for the alertness scale. **D.** A linear mixed effects model showed a significant interaction effect between group and time point for the fatigue scale. **E.** Visualization of the experimental protocol with regard to the time points at which the alertness and fatigue scales were administered.

1043 S.1.2 Determined TRF peak latency ranges

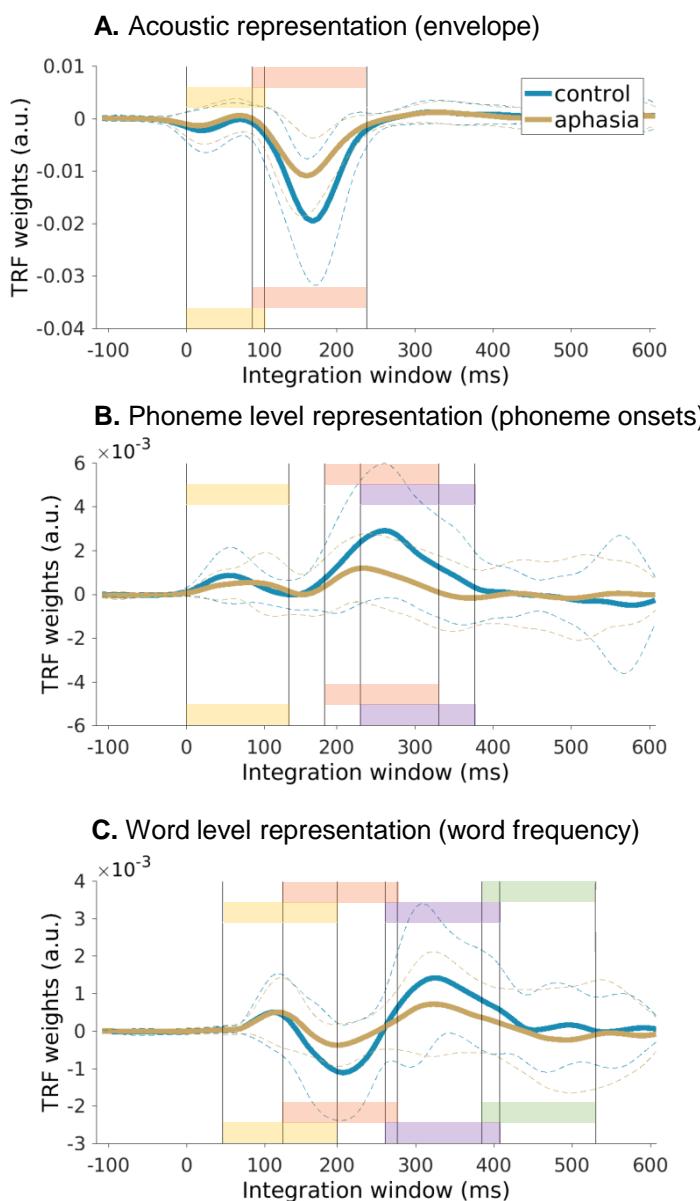


Figure S.5: Time ranges of 150 ms visualized in 3 speech representations.

1044 **S.1.3 Within-aphasia group exploratory correlations between prediction ac-**
1045 **curacy and behavioral measures**

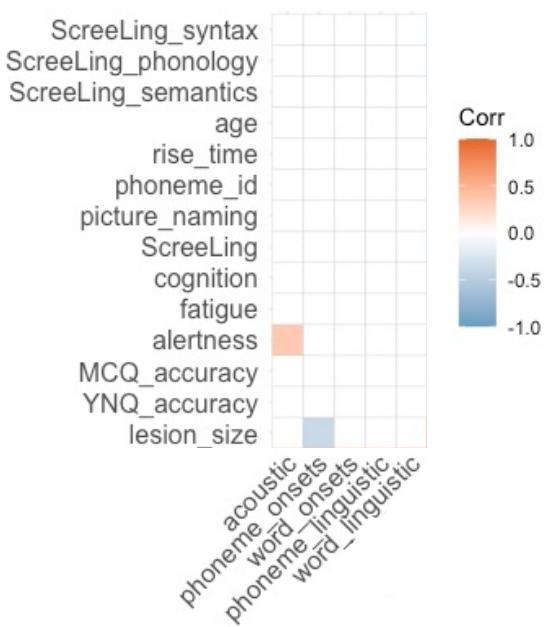


Figure S.6: **Exploratory correlation matrix.** Only significant correlations are colored. Note: Rise time and phoneme identification tasks are measuring acoustic and phoneme processing, the tasks are described in detail in (Kries et al., 2023). MCQ=multiple choice question; YNQ=Yes/No question.

- 1046 The phoneme onset prediction accuracy model is negatively correlated with lesion size: The higher the
1047 prediction accuracy, the smaller the size of the lesion.
1048 The acoustic prediction accuracy model is positively correlated with alertness: the higher the prediction
1049 accuracy, the higher the alertness.

1050 **S.1.4 Comparing the control group to a mild aphasia subgroup and an apha-**
1051 **sia subgroup with more severe impairment**

1052 The aphasia group in the main text consists of 39 IWA who all had language problems at least in
1053 the acute phase after stroke, as measured by scoring below the cut-off on a standardized diagnostic
1054 test (Comprehensive Aphasia Test-NL, ScreeLing, Aachen Aphasia Test). Aphasia is a disorder with a
1055 heterogeneous phenotype, as was indeed mirrored in the large variability of the language test outcomes
1056 (see table S.1). This made us wonder whether subtle effects may be hidden by group statistics due to
1057 this large variability. Therefore, we decided post-hoc to split up the aphasia group into a more mild and
1058 a more severe aphasia group. We based the splitting on the outcomes on the diagnostic test ScreeLing
1059 (max. score: 72; cut-off score: 68) and the picture-naming test NBT (max. score: 276; cut-off score:
1060 255) (fig. S.7), which resulted in a mild aphasia subgroup of 18 participants and an aphasia subgroup of

1061 21 participants. The control group consisted of 24 participants.

1062 Two inclusion criteria for the mild aphasia subgroup were formulated, i.e., not scoring below cut-off on
1063 either of the 2 tests at the time of data collection or scoring below cut-off on only one test, as long as
1064 the ScreeLing score is within the range of controls (range: 63-72). The reason to set this extra criterium
1065 for the ScreeLing, but not for the NBT, is that the ScreeLing is a diagnostic test characterizing language
1066 impairment across 12 different tasks that involve phonological, semantic and syntactic processing. Failing
1067 on the ScreeLing below the range of control participants is a clear indication of language impairment.
1068 The NBT on the other hand only tests one specific symptom of aphasia, i.e., anomia, and does not reflect
1069 a holistic assessment of language performance. Nonetheless, we included the NBT for categorizing the
1070 aphasia group, because anomia is present to varying extents across aphasia subtypes. Please note, the
1071 fact that IWA scored above cut-off threshold on both language tasks does not necessarily mean that they
1072 are recovered and do not have aphasia anymore. Many of these IWA were still in therapy at the moment
1073 of data collection and might have done the same tests at earlier times during therapy or might have
1074 learned compensation strategies in therapy. They had a left-hemispheric lesion after stroke and they had
1075 aphasia at least in the acute phase after stroke.

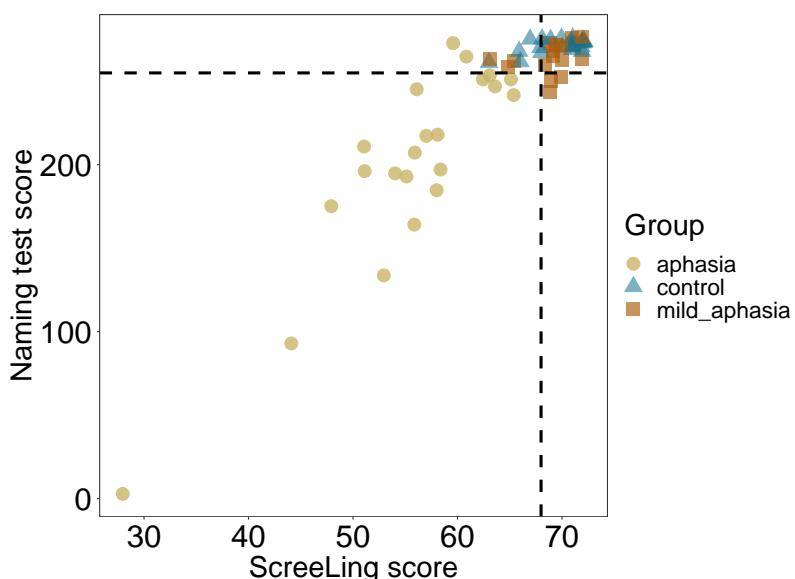


Figure S.7: Visualization of scores on language tests used for splitting up the aphasia group into a milder and a more severe group.

1076 We repeated the same analyses for these 3 groups (control, mild aphasia, aphasia) as for the 2 groups in
1077 the main text (control, total aphasia group), i.e., group comparisons of the model prediction accuracies
1078 averaged over all electrodes, cluster-based permutation tests of the topography of the prediction accu-
1079 racy and cluster-based permutation tests of the TRFs. These analyses were not corrected for multiple
1080 comparisons.

1081 S.1.4.1 Prediction accuracy

1082 We found a significant main effect of age for the acoustic model ($F(2) = 3.69$; $p = 0.031$) (fig. S.8
1083 A). Post-hoc pairwise comparisons revealed that there is a difference between the control and aphasia
1084 subgroup ($p = 0.032$), but not between the control and mild aphasia subgroup. We controlled for the
1085 influence of age, hearing, cognition, alertness and fatigue. None of the other neural tracking models
1086 showed any significant group effect.

1087 For the acoustic model, we found 2 significant clusters for the comparison between the control and aphasia
1088 subgroup (anterior left-lateralized cluster: $v = 32.728$, $p = 0.014$; posterior cluster: $v = 37.629$, $p = 0.009$),
1089 but none for the comparison between the control and mild aphasia subgroup (fig. S.8 B). The cluster
1090 locations are in line with the clusters found for the group comparison between the control group and the
1091 full aphasia group (fig. 4). None of the other models revealed any significant clusters between groups.

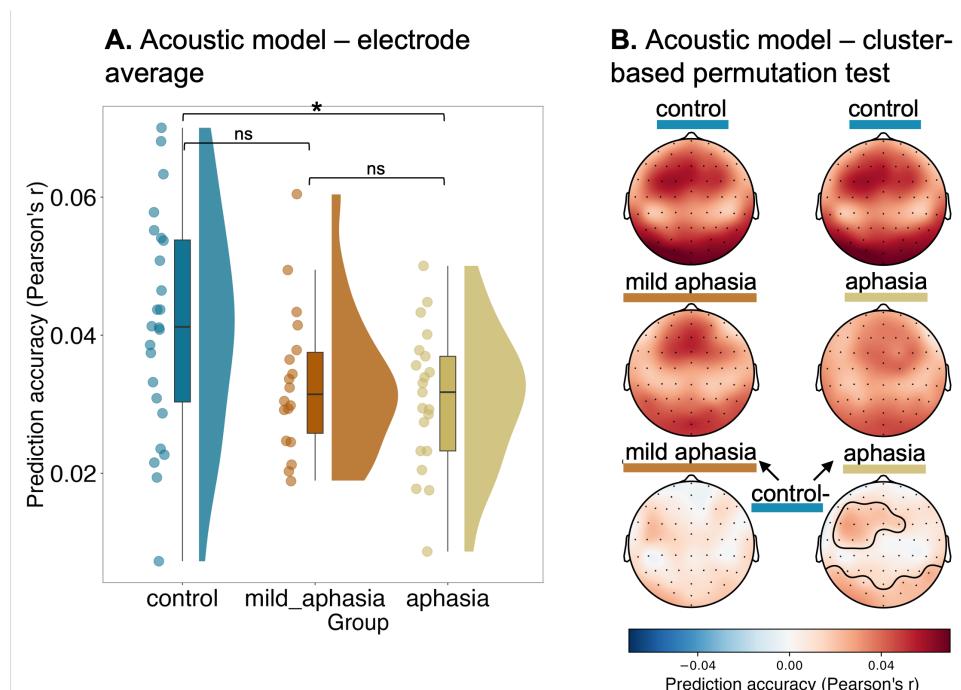


Figure S.8: **Only the subgroup of aphasia with more severe language difficulties shows decreased neural tracking of acoustic speech representations - across all electrodes and in local clusters.** **A.** When averaging the prediction accuracies of the acoustic model across all 64 electrodes, we found a significant group effect, even when we controlled for age, hearing, cognition, alertness, fatigue and lesion size. The acoustic model consists of the speech envelope and its onsets as speech representations. **B.** The cluster-based permutation test for the group comparison between control and aphasia subgroup, but not between the control and mild aphasia subgroup, revealed 2 clusters that significantly differed between groups, showing lower prediction accuracies in individuals with more severe aphasia. The lowest topoplot consists of the difference between the control group and the aphasia/mild aphasia group and the contours encircle the significant electrode clusters.

1092 S.1.4.2 Temporal response functions

1093 For the speech envelope TRF, we found 2 significant clusters for the comparison between control and
1094 aphasia subgroup around 180 milliseconds (ms) (anterior left-lateralized cluster: $v = 384.99$, $p = 0.007$;
1095 posterior cluster: $v = -271.52$, $p = 0.034$), but no significant clusters for the comparison between the
1096 control and the mild aphasia subgroup. The 2 significant clusters were located anteriorly left-lateralized
1097 and posteriorly, which is in line with the results from the TRF cluster-based permutation test between
1098 the control group and the full aphasia group in the main text (fig. 5), but also in line with the results
1099 from the prediction accuracy of the acoustic model (fig. 4 and S.8). Note that the time window of this
1100 cluster is also in line with the results in the main text (fig. 5).

1101 For word-level segmentation and linguistic speech representations, we also found differences between
1102 the control group and the aphasia subgroup, but not between the control group and the mild aphasia
1103 subgroup (fig. S.9). However, the time points of the significant clusters that were found are not the same
1104 as were found in the main text (5). In the main text (comparing the control group and the full aphasia
1105 group), we found clusters in the neural response around 200 ms for word onsets, word surprisal and word
1106 frequency, whereas for the comparison between the control group and the aphasia group with more severe
1107 difficulties, we found clusters in the neural response around 360 ms for the same speech representations
1108 (fig. S.9). For the neural response to all 3 representations, the clusters that differ between groups around
1109 360 ms are located over temporal left electrodes.

1110 For the other speech representations, no significant clusters were found.

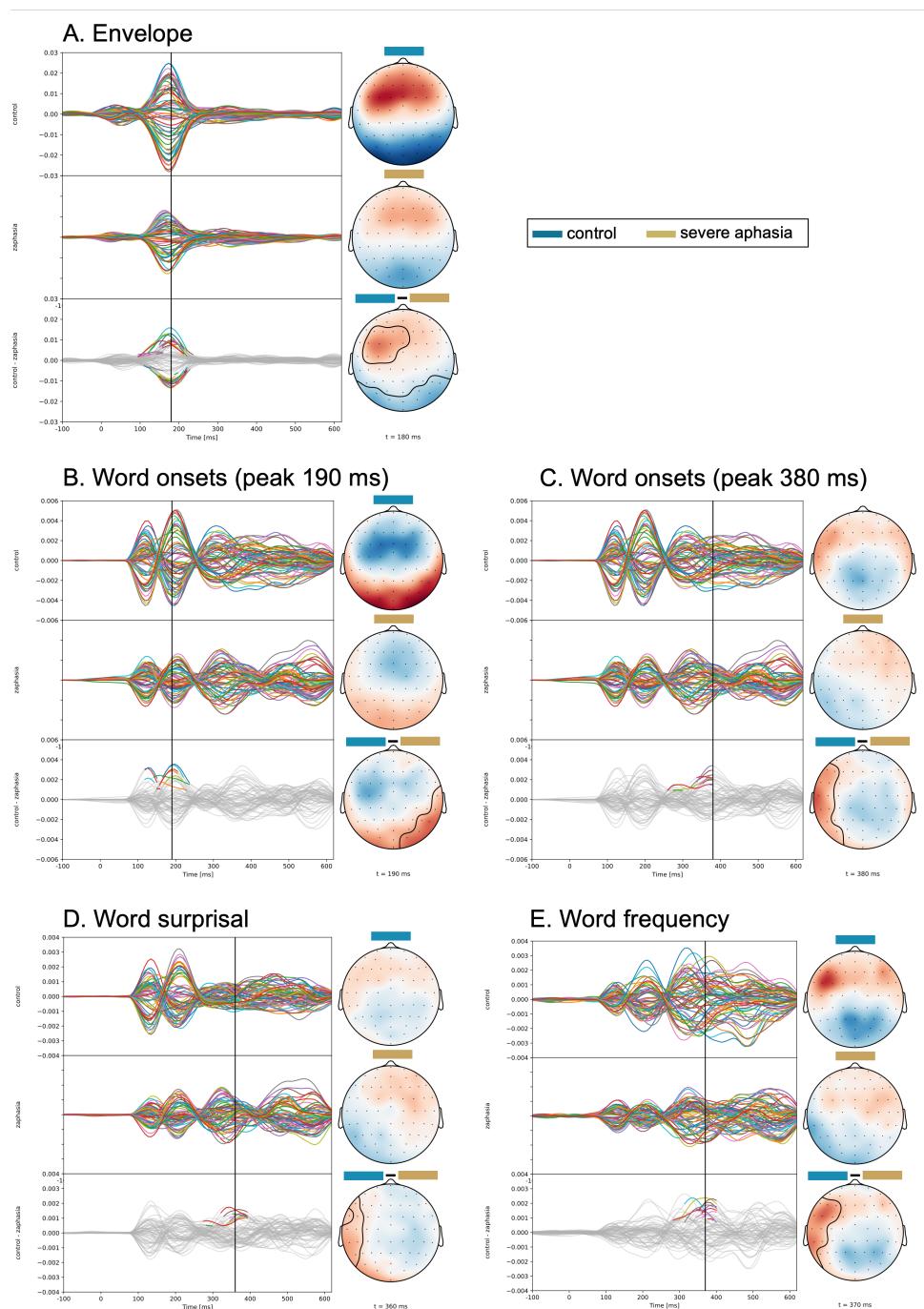


Figure S.9: Individuals with more severe aphasia show reduced neural response patterns to acoustic and linguistic representations in local clusters compared to controls. The top most topoplot of each panel shows the control group average and the middle plot shows the aphasia subgroup average. The spaghetti plots show the TRF weights across the integration window. Each line represents one of the 64 electrodes. The topoplots display the topography of the neural response at the time point indicated in the TRF plot with a vertical black line. The bottom most TRF plot and topoplot display the difference between the control group and the aphasia subgroup with more severe language problems. In the bottom TRF plot of each panel, the colored parts represent the electrodes within the significant cluster and the integration window over which they are significant.

1111 **S.1.4.3 Lesion volume in aphasia subgroups**

1112 A significant difference was found between the lesion volume in the mild aphasia group and the more
1113 severe aphasia group ($F(1, 37)=14.307$; $p=0.0005$; fig. S.10).

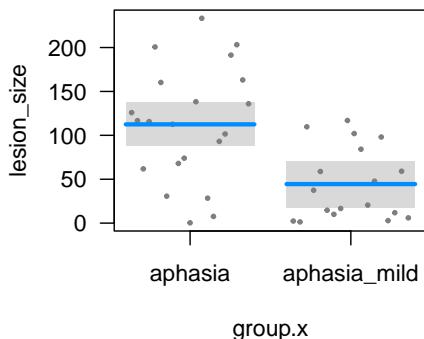


Figure S.10: **Lesion volume of the mild aphasia and more severe aphasia group.** The lesion volume is significantly larger in the more severe aphasia group as compared to the mild aphasia group.

1114 **S.1.5 Are IWA with a stroke origin in the posterior cerebral artery driving
1115 the posterior clusters?**

1116 Sub-010, sub-017 and sub-019 had a stroke originating in the posterior cerebral artery, which can affect
1117 occipital cortex, but also inferior/ middle temporal gyrus and parts of the parietal cortex. We were
1118 wondering whether the posterior clusters that we observe in the cluster-based permutation tests (figures
1119 4 and 5) may be driven by the 3 IWA who had a lesion originating from the posterior cerebral artery.
1120 Therefore, we repeated the cluster-based permutation tests for the acoustic model prediction accuracy
1121 and envelope TRFs without sub-010, sub-017 and sub-019.

1122 In figure S.11, we show the results. The clusters stay very similar as to when we conduct the tests with all
1123 39 IWA, and the posterior clusters remain present. The 3 left-out subjects do thus not drive the posterior
1124 clusters that we observe.

1125 When we conduct the group comparison on the electrode-average (with covariates, including lesion size;
1126 see section 2.4.1), the group difference also stays significant when the suggested 3 IWA are excluded
1127 ($F=6.198$; $p=0.016$).

1128 This supplementary analysis shows that the posterior clusters are not driven by the lesion site of the 3
1129 IWA who had a lesion originating from the posterior cerebral artery.

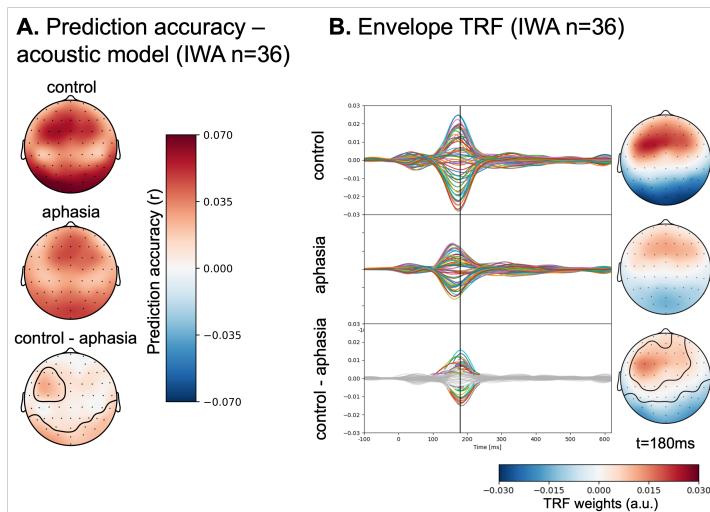


Figure S.11: The patterns resulting from the cluster-based permutations when the 3 IWA with lesions resulting from PCA damage are excluded are similar to the pattern of results when the whole aphasia group is compared to the control group. In the encircled clusters, the aphasia group had significantly lower amplitudes than the control group.