

1 **Phase-dependent word perception emerges from region-specific sensitivity to**
2 **the statistics of language**

3

4 Sanne Ten Oever^{1,2,3*}, Lorenzo Titone⁴, Noémie te Rietmolen^{1,2}, Andrea E. Martin^{1,2}

5

6 ¹ Language and Computation in Neural Systems group, Max Planck Institute for Psycholinguistics, Nijmegen, The
7 Netherlands

8 ² Language and Computation in Neural Systems group, Donders Centre for Cognitive Neuroimaging, Nijmegen, The
9 Netherlands

10 ³ Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, The
11 Netherlands

12 ⁴ Research Group Language Cycles, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

13

14 * Corresponding author: sanne.tenoever@maastrichtuniversity.nl

15 **Abstract**

16 Neural oscillations reflect fluctuations in excitability, which biases the percept of ambiguous sensory
17 input. Why this bias occurs is still not fully understood. We hypothesized that neural populations
18 representing likely events are more sensitive, and thereby become active on earlier oscillatory phases,
19 when the ensemble itself is less excitable. Perception of ambiguous input presented during less-excitable
20 phases should therefore be biased towards frequent or predictable stimuli that have lower activation
21 thresholds. Here, we show with computational modelling, psychophysics, and magnetoencephalography
22 such a frequency bias in spoken word recognition; a computational model matched the double
23 dissociation found with MEG, where the phase of oscillations in the superior temporal gyrus (STG) and
24 medial temporal gyrus (MTG) biased word-identification behavior based on phoneme and lexical
25 frequencies, respectively. These results demonstrate that oscillations provide a temporal ordering of
26 neural activity based on the sensitivity of separable neural populations.

27 Introduction

28 Oscillations, or population rhythmic activity, reflect the waxing and waning of neural excitability such that
29 individual neurons are modulated by an oscillation are primarily active on high excitable phases¹⁻³.
30 Previous studies have directly linked this phase-dependent neural activity to behavioral performance
31 showing that target detection⁴⁻⁷ and accuracy in categorization tasks^{8,9} are modulated by oscillatory
32 phase. Besides accuracy, a few studies have also shown that oscillatory phase can modulate the
33 categorization of ambiguous stimuli by biasing participants' percept to one or another category based on
34 the phase of presentation¹⁰⁻¹². Improved behavioral performance has often been attributed to increased
35 processing efficiency at oscillatory phases at which neural activity is increased^{3,13}. However, phase-
36 dependent categorization biases cannot be explained by overall increases in activity (or increased
37 processing efficiency) on specific oscillatory phases because increases in overall activity should not bias
38 processing to one specific perceptual interpretation. Thus, it is unclear what neural mechanism underlies
39 phase-dependent categorization.

40 Even though oscillations modulate neural excitability, not all neurons influenced by an oscillation
41 reach activation exactly at the same time or phase. In fact, the phase-of-firing of a neuron is determined
42 by an interaction between excitability changes due to oscillations and the neural sensitivity of a neuron
43 to incoming signals¹⁴. Neurons that are sensitive to incoming signals will respond strongly and will
44 therefore reach activation already at relatively low excitable oscillatory phases^{1,14}. In contrast, neurons
45 less sensitive to the input will reach activation only at later, more excitable, phases. Neuronal sensitivity
46 can be modulated by neuro-plastic changes induced through associative and statistical learning¹⁵. For
47 example, neural populations representing more likely events in the world have higher sensitivity than
48 populations representing less likely events^{15,16}. If this is true, populations representing probable events
49 should be active at earlier, less excitable, oscillatory phases compared to populations representing less
50 likely events which in turn could lead to phase-dependent categorization¹⁷.

51 Previously, we have shown that oscillatory phase in the theta frequency range can bias the
52 categorization of an ambiguous syllable¹¹. In that study, we presented an ambiguous syllable that Dutch
53 participants could interpret as /da/ or /xa/ (notation according to the international phonetic alphabet
54 [IPA]). Originally, this phase-dependent categorization bias was attributed to an articulatory visual-to-
55 auditory temporal difference between the two syllables (the visual-to-auditory articulatory delay of /da/
56 is shorter than /xa/)^{11,18}. However, in Dutch, /d/ also has a higher frequency than /x/¹⁹, that is, /d/ is more
57 probable than /x/. It is therefore possible that the categorization effect in this study was instead (or

Table 1. Stimulus materials

Dutch word	<i>gat</i>	<i>dat</i>	<i>gaat</i>	<i>daad</i>
Translation	<i>hole</i>	<i>that</i>	<i>go</i>	<i>deed</i>
IPA	/xat/	/dat/	/xat/	/dat/
consonant frequency	--	++	--	++
vowel frequency	++	++	--	--
word frequency	--	++	++	--
frequency denotation	cVw	CVW	cvW	Cvw

Four different Dutch words used in the study. Plusses and minuses indicate whether the trait was frequent (++) or non-frequent (--). IPA = international phonetic alphabet. In the rest of the text, we will denote the frequency across the three traits (consonant (C), vowel (V), and word (W)), by indicating via lower- or upper-case letters whether the word has a low (lower-case) or high (upper-case) frequency for a specific trait (see denotation at the last row).

58 additionally) caused by an interaction between ongoing oscillations and the sensitivity of neural
59 populations for these consonants¹⁷. Up to now, no study has investigated the role of event probability for
60 the outcome of phase-dependent behavior. If oscillations provide a temporal ordering based on neural
61 sensitivity^{14,17}, then one would expect phasic categorization effects to occur when there is a difference in
62 event probability for two possible interpretations of an ambiguous stimulus. If this is so, we should view
63 oscillations not merely as a gating operation opening and closing lines of neural communication^{3,13}, but
64 rather as a rich source of representational space^{20,21}.

65 To investigate the relation between phase-dependent categorization and event probability we
66 presented participants with words that varied in consonant, vowel, and word frequency. These were the
67 four Dutch words *dat*, *gat*, *daad*, and *gaat* (see table 1 for translations, IPA notation, and event
68 probability/frequencies). In this way, we manipulated event probabilities at different levels of analysis
69 based on the recurrence of phonological and lexical elements in a language. By using computational
70 modelling, psychophysics, and MEG we could investigate the influence of event probability on behavioral
71 and neural responses to ambiguous stimuli (Figure 1). Computational modelling showed that phase biases
72 categorization when words had different probabilities of occurrence. Moreover, the bias was dependent
73 on the event probability within the specific level of analysis (phoneme or word level). These outcomes
74 were verified using psychophysics and MEG.

75

76

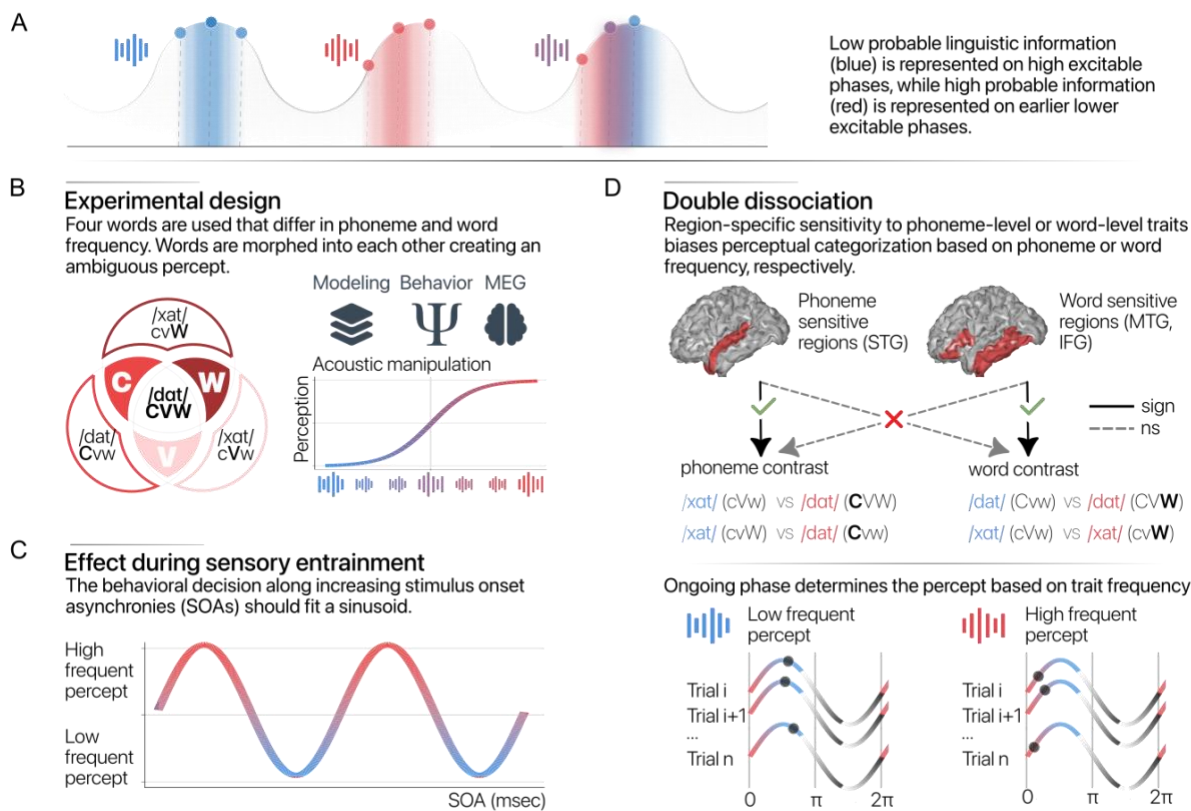


Figure 1. Overview of the current study. A) It is hypothesized that low probable linguistic information is represented at high excitable phases while high probable linguistic information is represented at low excitable phases. The perceived sound of ambiguous stimuli depends on the phase of presentation. B) Four words are used that differ in consonant (C), vowel (V), and word (W) frequency. Words are morphed into each other creating an ambiguous percept. C) Sensory entrainment locks neural rhythms to the rhythmic input and therefore the stimulus onset asynchrony (SOA) relative to the entrainment train should bias the percept to words containing low or high frequent linguistic information. D) In MEG a double dissociation is expected in which perceptual bias is governed by phoneme or word frequency for regions sensitive to phonemic or word features, respectively. Sign = significant; ns = not significant. STG = superior temporal gyrus; MTG = middle temporal gyrus; IFG = inferior frontal gyrus.

77 Results

78 In the current study, we investigated whether combining oscillations with dissimilar ‘event’ probabilities
79 – here phoneme and lexical frequency - can lead to phase-dependent categorization. First, we designed a
80 computational model whose goal was to categorize incoming sensory input. To directly link the outcome
81 of the computational model to empirical findings and quantify phase-dependent categorization, we
82 presented the model with the stimuli used in the psychophysics and MEG experiment. In the
83 psychophysics study, we presented an entrainment stream at 6.25 Hz after which an ambiguous word was
84 presented at a variable stimulus onset asynchronies (SOAs). Assuming an entrained oscillation, these SOAs
85 match to ongoing oscillatory phases²². We generated ambiguous words by creating 10 equally spaced
86 morph levels along each dimension (that is varying either the consonant /x/-/d/ or the vowel /a/-/a/). This

87 procedure resulted in four different morph scales: /xat/-/dat/, /xat/-/dat/, /dat/-/dat/, and /xat/-/xat/.
88 Participants categorized each morphed stimulus as one of either word. The most ambiguous morph was
89 defined individually for each participant by fitting a psychometric curve to their responses and selecting
90 the morph closest to 50% categorization. In the MEG experiment, stimuli were not preceded with an
91 entrainment train, but were presented at random SOAs.

92

93 **Computational model**

94 The computational model was an extension of the Speech Tracking in a Model Constrained Oscillatory
95 Network (STIMCON) model introduced in ¹⁷. This model integrates temporal tracking together with the
96 tracking of the content of speech using both temporal and content information. Input activation levels at
97 a given time ($A_{l,T}$) are governed by the following function:

$$98 \quad A_{l,T} = C_{l-1 \rightarrow l} * A_{l-1,T} + \text{inhib}(Ta) + \text{osc}(T) \quad (1)$$

99 in which C represents the connectivity patterns between different hierarchical levels (l), T the time in
100 milliseconds, and Ta a vector representing the times of individual nodes within the inhibition function (see
101 online methods). Input activation is thus determined by activations from lower levels as well as an
102 inhibition and an oscillation function. Individual Ta node values are set to zero as soon as activation of a
103 node reaches activation threshold (default threshold = 1). This function first ensures non-linear supra-
104 threshold activation after which the node is temporally inhibited.

105 In the current implementation, we introduced two different levels of analysis: a phoneme and a
106 word level. Both levels receive input from the input level but have their own connectivity with the input
107 and their own node sensitivity. The input is modelled as the individual words: /xat/, /xat/, /dat/, /dat/,
108 and an empty word node is used for the entrainment train. For the phoneme level, we represented the
109 phonemes /x/, /d/, /a/, /α/, and an empty phoneme node. Connectivity for the input-to-phoneme level
110 was set to one when the phoneme was part of the input word (the entrainment stimulus was connected
111 with a one to the empty phoneme node). The input-to-word level connectivity consists of an identity
112 matrix (each word loads with one on the word level). Sensitivity of individual nodes to input was varied
113 by reducing the activation threshold for the more frequent phonemes (/d/ and /α/) and words (/dat/ and
114 /gat/) in their respective analysis level (activation thresholds were parametrically reduced between 0.01
115 - 0.1). In all simulations, we extracted the categorization response of the model by determining the
116 deciding node that was active first after stimulus presentation. For the two categorization options, we

117 coded one node as a 0 and the other node as a 1 (if both nodes were active simultaneously or no node
 118 was activated at all we assumed that the model would guess and set the value to 0.5 for the psychophysics
 119 simulation or choose randomly 0 or 1 for the MEG simulation). For the psychophysics experiment, we
 120 assumed that the output of the model reflects the average outcome of the phoneme and word level.

121 We let the model run through the psychophysics and MEG experiments. In both experiments, all
 122 morphs are initially presented at random moments to generate a psychometric curve and to determine
 123 the most ambiguous stimulus that will be used for the main experiment (see online methods for more

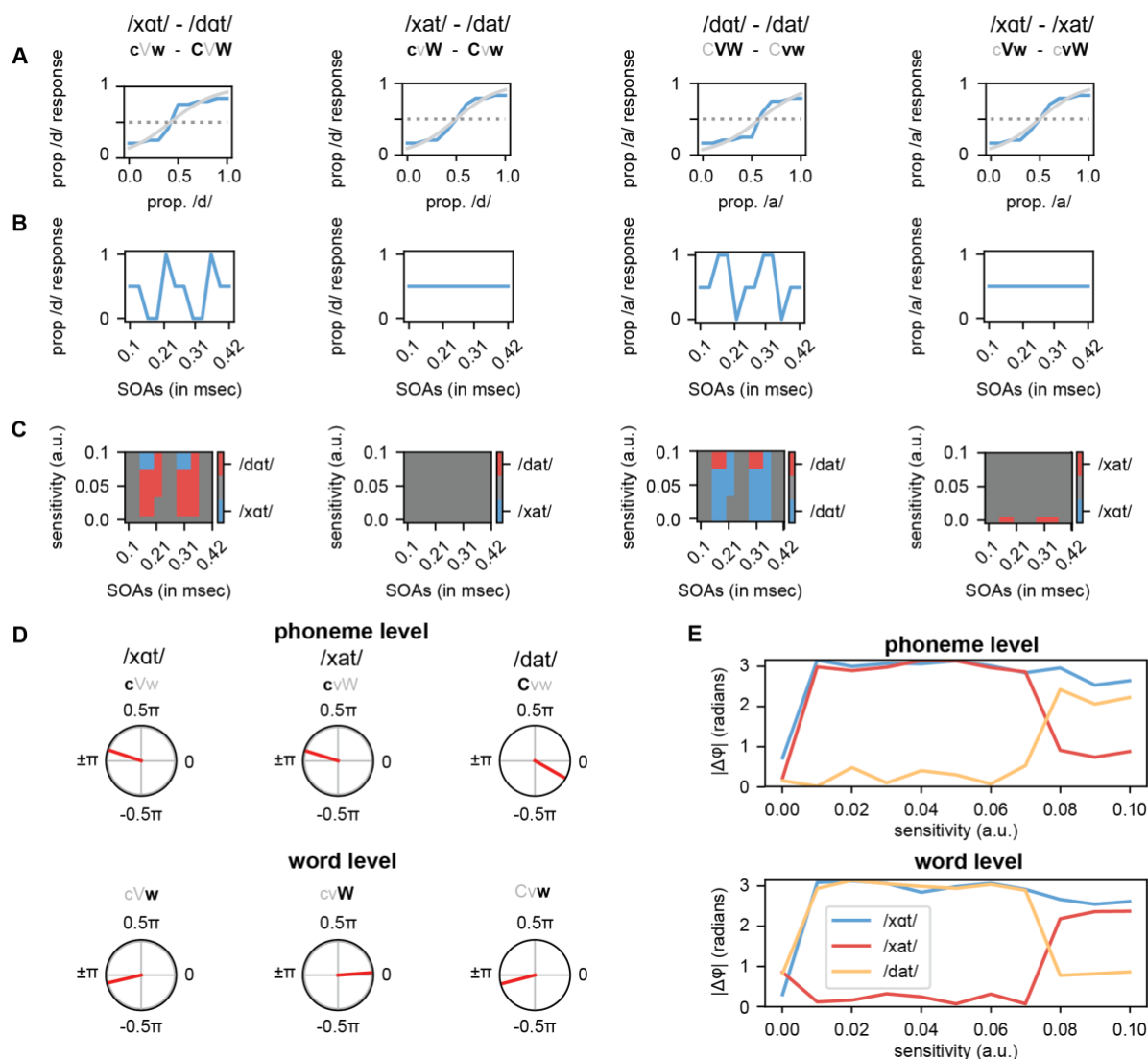


Figure 2. Outcome of the computational model. A) Psychometric functions for the four different morph dimensions (sensitivity = 0.09). Blue lines represent the model output; grey lines show the psychometric fit. B) Response of the model to the most ambiguous morph of A presented at different stimulus onset asynchronies. C) Response choice of the model during entrainment (as in B) for different sensitivity levels. D) Phase difference between the average phase of the three different response choices and /dat/ (sensitivity = 0.07). E) Phase differences (as in D) for different sensitivity levels.

124 details). To imitate this procedure, input was presented at different amplitude proportions of two words
125 (e.g. for /xat/-/dat/ morph with 90% of /dat/, /xat/ was presented with an amplitude of 0.1 and /dat/ at
126 0.9) evenly distributed across all phase values. For each morph, we averaged the node responses of the
127 model across the repetitions of the same morph (and across the two levels). For all word morph spectra,
128 we could reliably fit a psychometric function and extract the most ambiguous morph (Figure 2A). For the
129 second part of the psychophysics experiment, the model was presented with an entrainment train of
130 empty words after which we presented the most ambiguous morph stimulus at different SOAs. Results
131 show that only for morph spectra in which the two morphed traits had opposing event probabilities or
132 frequency (i.e. /dat/ and /xat/ [CVW vs cVw] and /dat/ and /dat/ [CVW vs Cvw]) a phase-dependent
133 categorization performance developed across all sensitivities tested (Figure 2B and 2C).

134 For the MEG experiment, phase does not have to be inferred from an entrainment train, but
135 rather can be estimated from the recorded regions directly. To simulate this experiment, ambiguous
136 morphs were presented to the model at random phases (repeated for 1000 repetitions). Only the
137 consonant morphs along the /dat/-/xat/ and /dat/-/xat/ spectra were used in the MEG experiment. For
138 each of the two ambiguous morphs, the phase was extracted together with the categorization response
139 based on the node activation of the phoneme or word level. For the main MEG experiment, we are not
140 hypothesizing about the absolute phase, as we have no hypothesis about the exact phase (see^{6,11}), but
141 rather in *relative* phase difference between a more-likely versus less-likely event. Therefore, we took the
142 phase difference between the word /dat/ which has a high frequency on all trait dimensions (CVW) and
143 the phase of the other response options. For the phoneme level, the model showed high phase differences
144 of around π when the ambiguous morph was interpreted as either /xat/ or /xat/, but low phase
145 differences of around 0 when the model interpreted the morph as /dat/ (Figure 2D). In contrast, for the
146 word level, we found π phase differences for the categorization choices /xat/ and /dat/, and 0 phase
147 difference for the choice /xat/. Thus, phase differences were low when both words had high frequency
148 traits within the level of analysis (high frequency phonemes in the phoneme level [/dat/, CVW vs /dat/,
149 Cvw] and high frequency words in the word level [/dat/, CVW vs /xat/, cvW]). Phase differences were high
150 when the words had different frequency traits (high versus low frequency phonemes in the phoneme level
151 [/dat/, CVW vs /xat/ [cVw] and /xat/, [cvW] and high versus low frequency words in the word level [/dat/,
152 CVW vs /xat/ [cVW] and /dat/, [Cvw]). In MEG, this pattern of results could correspond to phase
153 differences in different neural sources that analyze phoneme- and word-level representations,
154 respectively. These phase differences were more pronounced when the sensitivity level changes were
155 relatively low (Figure 2E).

156

157 Psychophysics experiment

158 We conducted a consonant and a vowel version of the experiment by morphing either the consonants of
 159 the words or the vowels of the words, creating four morphs: /xat/-/dat/, /xat/-/dat/, /dat/-/dat/, and
 160 /xat/-/xat/. During the first part of the experiment, we presented stimuli across all morphs to create an
 161 individual psychometric curve along the consonant (Figure 3A) and vowel dimension (Figure 3D). Only
 162 participants for which we could reliably extract an ambiguous stimulus via fitting a psychometric curve
 163 could participate in the main experiment.

164 In the main experiment, participants listened to rhythmic broadband noise bursts presented at
 165 6.25 Hz after which an ambiguous word was presented. The SOAs at which ambiguous words were
 166 presented ranged between 0.1 and 0.4 seconds in 12 equidistant steps (spanning exactly two cycles of
 167 6.25 Hz). Participants had to indicate which word they heard. For the consonant experiment, we found
 168 that a 6.25 Hz sinusoid fitted to the data yielded a higher explained variance than expected by chance for
 169 the /xat/-/dat/ morph ($p = 0.006$, $r^2 = 0.661$; Figure 3C), but not for the /xat/-/dat/ morph ($p > 0.05$; Figure
 170 3B). For the vowel experiment, we could significantly fit a sinusoid for the /dat/-/dat/ ($p = 0.039$, $r^2 =$

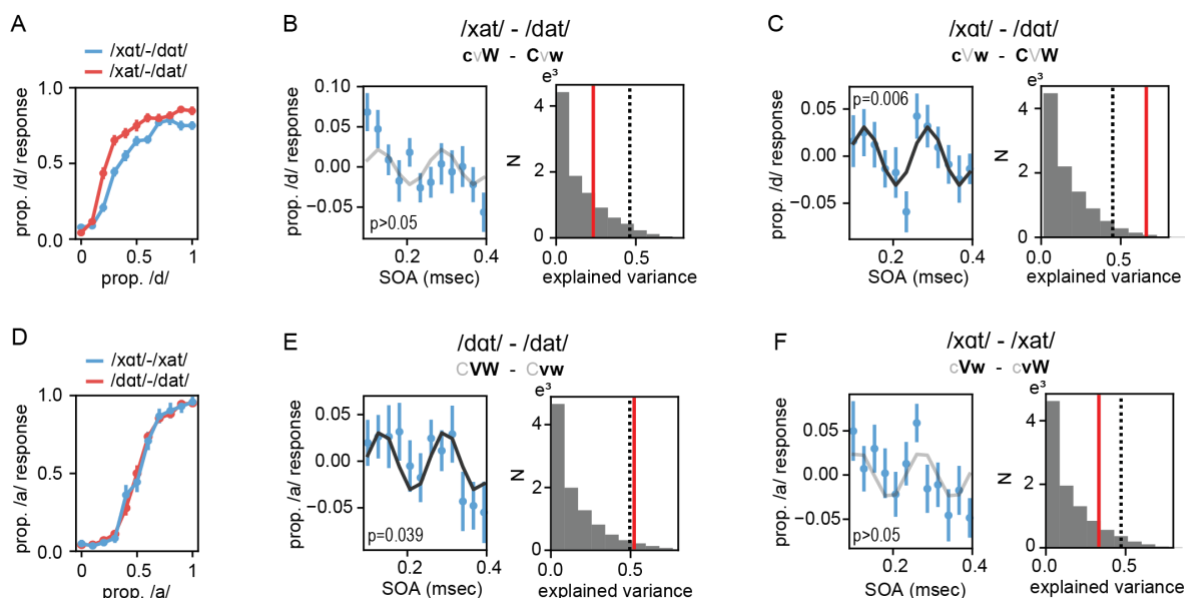


Figure 3. Behavioral results. A) Psychometric curves for the consonant experiment (separate lines for the two vowel types). B) Outcome of the main experiment for the /xat/-/dat/ spectrum. The left panel shows the average demeaned time course across stimulus onset asynchronies (SOA). The right panel shows a histogram of the null distribution together with the observed explained variance of the sinusoid fit (red line) and the 95th percentile (dotted line) of the null distribution. C) Same as B for the /xat/-/dat/ spectrum. D-F) Same as A-C for the vowel experiment. Error bars indicate the standard error of the mean. Black lines indicate the best fitted curve (gray if not significant).

171 0.523; Figure 3E), but not the /xat/-/xat/ morph ($p > 0.05$; Figure 3F). In sum, we could only fit a significant
172 curve for morphs in which both varied traits had opposing frequencies in the word pairs, that is for the
173 /xat/ - /dat/, cVw – CVW, morph and the /dat/ - /dat/, CVW – Cvw, morph. This was in line with the
174 outcome of the computational model (Figure 2).

175

176 MEG-experiment

177 The output of our computational model showed that event probability determines whether we can expect
178 phase-dependent categorization. Moreover, the phase-dependent categorization was different for the
179 phoneme and the word level of analysis. However, in the psychophysics experiment, all responses are
180 based on an integration across both levels of analyses as there is only one behavioral output. To bridge
181 this gap, we designed a MEG study in which the phase-dependent categorization effects at different levels
182 of analysis could be source localized to cortical regions that are known to correspond to varying levels of
183 linguistic computations^{23,24}. Indeed, earlier auditory regions such as superior temporal gyrus (STG) are
184 more sensitive to phoneme content, such as vowel and consonant traits, while regions higher in the
185 auditory hierarchy, such as medial temporal gyrus (MTG) and inferior frontal gyrus (IFG)
186 are sensitive to lexical representation and temporal integration, respectively^{23,25-28}. If
187 phoneme or word frequency relate to phase-dependent categorization, the phase
188 of ongoing oscillations in distinct cortical
189 regions should bias categorization based on
190 the level of analysis of that region. To test
191 this hypothesis, we presented the
192 ambiguous morphs of the consonant
193 experiment to Dutch participants while
194 recording their neural activity with MEG and
195 source localizing the response to the STG,
196 MTG and IFG.

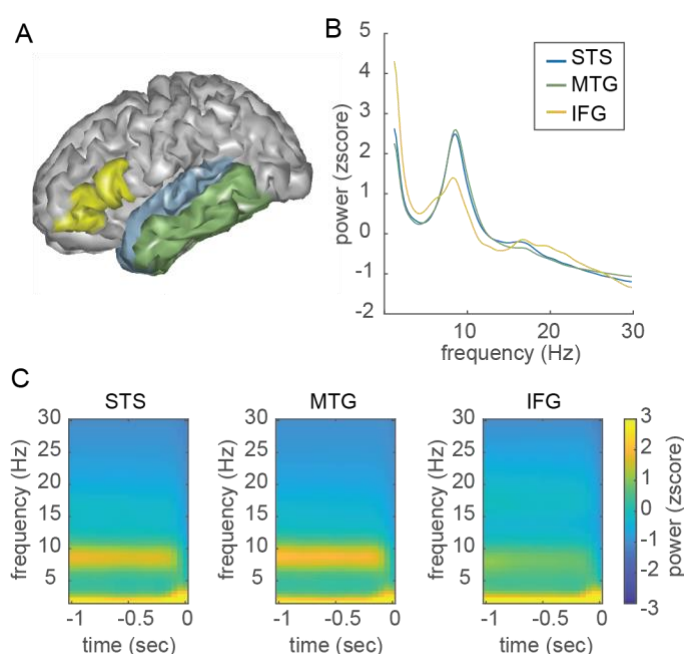


Figure 4. Pre-stimulus power in regions of interest. A) The three regions of interest. B) Power spectra averaged for the -1-0 sec time window averaged across the two ambiguous stimuli and response choices. C) Time-frequency response averaged across the ambiguous stimuli and response options. Note that data is padded from time point 0 on (explaining the sharp power drop around zero).

200 First we looked at the overall power response in the pre-stimulus period (see Supplementary
201 Figure 1 for post-stimulus responses). All analyses were based on using an array-gain beamformer which
202 corrects for center-of-head biases without the need of a baseline^{29,30}. To limit computational resources,
203 we focused on the first PCA of all grid points part of the ROI quantifying the component explaining the
204 most variance in the ongoing data. In all ROIs we found theta peak (peak values: 8.4 Hz, 8.6 Hz, and 8.2
205 Hz for STS, MTG, and IFG, respectively) across the whole pre-stimulus window, but this peak was weaker
206 for the IFG (Figure 4). We compared pre-stimulus power values dependent on the response of the
207 participant for the ambiguous stimuli but found no differences (all $p > 0.573$).

208 For all participants, we could individually determine the most ambiguous morph in the first part
209 of the experiment and all, but one could maintain an ambiguous percept throughout the experiment
210 (Figure 5A-B). We investigated whether phase-dependent categorization was determined by phoneme or
211 word frequency for each ROI separately. At each pre-stimulus time-frequency point, we performed a
212 circular-linear correlation between the pre-stimulus phase and response type (low or high trait frequency;
213 that is, for the consonant contrast /xat/ [cVw] and /xat/ [cvW] vs /dat/ [CVW] and /dat/ [Cvw]; for the
214 word contrast /xat/ [cVw] and /dat/ [Cvw] vs /xat/ [cvW] and /dat/ [CVW]). To correct for multiple
215 comparisons, we ran cluster-based statistics³¹. For the consonant contrast we found a significant effect of
216 consonant frequency in the STG (Figure 5C; cluster statistic: 69.655; p-value: 0.022; frequency range: 4.4
217 - 9.0 Hz; time range: -0.30 - -0.10 sec; peak $t(21)$ -value: 4.119 at 7.617 Hz, -0.20 sec), but not in the MTG
218 or IFG ($p > 0.05$). For the word contrast, we found a significant effect of word frequency in the MTG (Figure
219 5C; cluster statistic: 66.138, p-value: 0.019; frequency range: 4.4 - 8.5 Hz; time range: -0.25 - 0 sec; peak
220 $t(21)$ -value: 3.589 at 6.445 Hz, -0.10 sec), but not in the STG or IFG ($p > 0.05$). In sum, we found a double
221 dissociation between ROI and trait type.

222 To further evaluate the exact phase differences for each ambiguous sound, we computed the
223 average phase at which participants heard either of the two words. Phases were extracted for each
224 individual's peak time-frequency point within the significant cluster (for both morphs separately). The
225 exact MEG phase was not expected to be identical across participants as it is difficult to determine
226 excitability levels of an oscillation from MEG phase and individual stimulus processing times might differ.
227 Rather, high event probabilities should be represented at the same phase, while low event probabilities
228 are represented at the opposite phase within each participant. We took the word /dat/, which had high
229 frequencies on all traits (CVW word), as a reference word and took the phase difference between the
230 average phase at which participants heard /dat/ and one of the three words (ambiguous words perceived

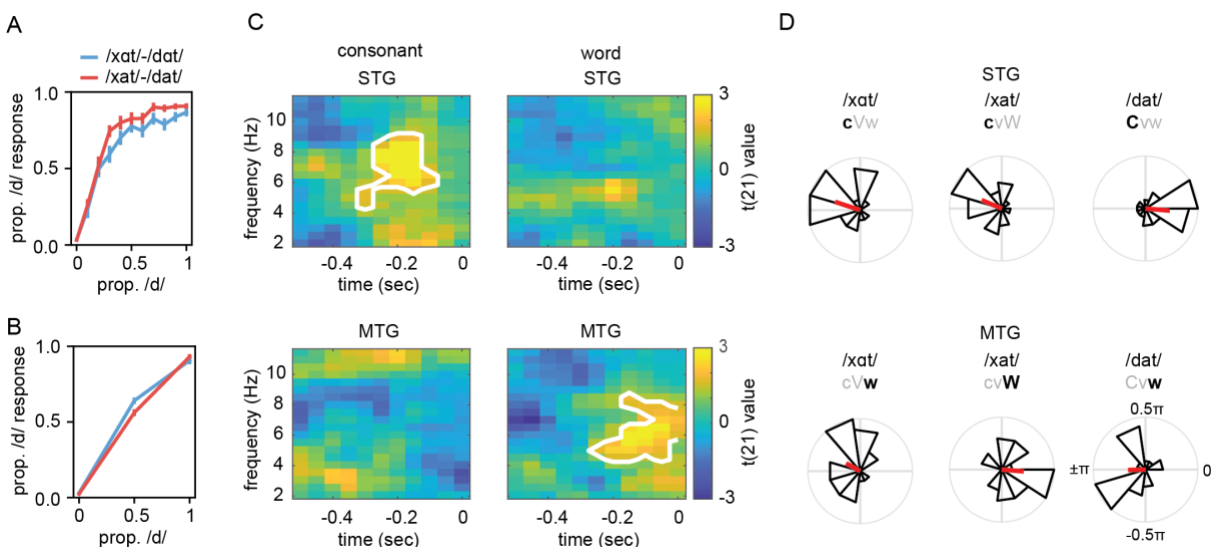


Figure 5. MEG results. A) Behavioral responses in the MEG experiment for part one of the experiment. B) Behavioral response for the main experiment. Error bars reflect the standard error of the mean. C) Circular-linear correlation results show that in STG phase determines the percept of consonant frequency (top) while in MTG phase determines the percept of word frequency (bottom). White outline indicates significance ($p < 0.05$) according to the cluster-based statistical test. D) Phase difference with CVW /dat/ at the individual peak time-frequency point in C. In STG, words with high frequency consonants (Cxx words) had a phase difference of 0, while words with low frequency consonants (cxw words) had a phase difference of around π (top). In MTG, high frequency words (xxW words) had a phase difference of 0, while low frequency words (xxw words) had a phase difference of around π (bottom).

231 as /xat/, /xat/, or /dat/). We expected that for high frequency traits the phase difference would be zero,
 232 while for low frequency traits the phase difference would be around π relative to the reference word. To
 233 test this, we performed a v-test that tests for non-uniformity with a specific direction for each of the three
 234 words per ROI. To generate a p-value that combines the three values (as we expected all three contrasts
 235 to be significant), we multiplied the three probabilities yielding the probability of all three events
 236 happening at the same time (assuming independent tests).

237 In STG, we found that the average phase difference for the three words /xat/, /xat/, or /dat/ was
 238 0.90π , 0.87π , and -0.027π respectively. The three phase differences were close to the expected phase
 239 differences: the low frequency consonants having a phase difference of π , while the high frequency
 240 consonant having a phase difference of zero (Figure 5D). The individual tests showed significance (/xat/
 241 vstat = 10.56, pval < 0.001; /xat/: vstat = 8.47, pval = 0.005; /dat/: vstat = 10.18, pval = 0.001 respectively)
 242 as well as the combined probability ($p < 0.001$). In MTG, the average phase difference for the three words
 243 /xat/, /xat/, or /dat/ was 0.84π , -0.026π , and -0.99π respectively. The phase differences were close to the
 244 expected phase differences: the low frequency words having a phase difference of π and the high
 245 frequency word having a phase difference of zero. The individual tests showed significance (/xat/: vstat =

246 5.91.56, pval = 0.037; /xat/: vstat = 9.26, pval = 0.003; /dat/: vstat = 7.36, pval = 0.013 respectively) as well
247 as the combined probability ($p < 0.001$). Thus, also for the individual words, the phase differences matched
248 the neural sensitivity of the underlying region.

249 Discussion

250 In the current study we used computational modelling, psychophysics, and MEG recordings to
251 demonstrate that variations in neural sensitivity to event probability operationalized as phoneme and
252 word frequency can result in phase-dependent perceptual categorization. We showed that ambiguous
253 words presented at different phases, either through neural entrainment or by extracting the phase from
254 MEG, are interpreted as one or another word depending on the time or phase of presentation. Moreover,
255 in the MEG data we could dissociate these effects to separate cortical regions: phase-dependent
256 categorization in STG depended on phoneme frequency, while word frequency modulated phase-
257 dependent responses in MTG. The behavioral findings and the double dissociation between STG and MTG
258 responses matched the results from a computational model that uses oscillations, together with varying
259 neural sensitivity, to capture categorization responses. These results demonstrate that the neural phase
260 code relies on ordering based on neural sensitivity, directly linking phase-coding to behavioral outcomes
261 in a categorization task.

262 Most studies investigating the direct link between ongoing oscillatory phase and behavior have
263 focused on assessing the role of oscillatory phase in modulating overall performance measures, such as
264 accuracy^{8,32}, detection^{4,5,33}, or reaction times³⁴. These studies are all based on the assumption that
265 oscillations modulate overall firing rates and subsequent neural processing should be optimized at phases
266 where neural excitability is high³. We here argue that this view might be too simplistic and does not
267 provide the full picture of the role of oscillations for neural computation. Instead of merely providing
268 windows of processing opportunity^{3,13}, oscillations provide a means to organize the complex neural
269 dynamics by activating neural populations at different neural phases^{14,20,35}. This organizational principle
270 of phase coding has extensively been shown with invasive recordings in animals^{36,37}. Moreover, prior
271 computational modelling has shown the computational benefit of this organizational principle, as it
272 effectively increases the representational space in the brain³⁵ and changes the formal expressive power
273 of a system³⁸⁻⁴⁰.

274 It has been an open debate what are the organizational principles of phase coding, that is, what
275 determines when a neuron is activated³⁷. In working memory paradigms, sequence order has often been
276 implied to be the main organizational principle of phase coding^{41,42}. This is based on studies primarily in
277 rats that show phase precession in which the order of the phase code is linked to the order of upcoming
278 locations in an explorative maze task³⁶. In our study, no sequence order can be imposed. Nonetheless, we
279 find that the behavioral outcome of a phase code is based on the frequency of phonemes and words in

280 the Dutch language by using computational modeling, psychophysics, and MEG. This finding suggests that
281 sequence order is not the only principle that can generate a phase code. Instead, overall event probability
282 of words within a language modulates the neural sensitivity and can also impose a phase code¹⁷. While
283 the present study is focused on event probability based on the overall frequency of information in a
284 language, we hypothesize that this finding can be extended to event probability that also depends on
285 contextual knowledge. Evidently, contextual event probability and sequential order are related: events
286 that are going to happen earlier in the future are more probable in the short-term. Moreover, events
287 occurring in the near future have a higher behavioral relevance. Both probability and behavioral relevance
288 could have a consequence for how excitable individual neural populations are. Therefore, excitability
289 shifts, rather than solely order or event probability, could be the core principle that organizes the phase
290 code.

291 The brain's sensitivity to varying levels of event likelihood based on the statistics in the world has
292 been shown in a plethora of studies which demonstrate that the brain is more sensitive to stimuli that are
293 more likely¹⁶. However, the consequence of a probability manipulation in combination with oscillatory
294 dynamics has rarely been studied. We here provide the first behavioral evidence showing that event
295 probability and oscillations together provide a phase code which activates event representations on
296 different phases based on their likelihood. It is unknown how strongly event statistics modulate neural
297 sensitivity. This could potentially be relevant for the behavior of a neural system as our computational
298 model suggests a non-linear change in phase-dependent perceptual outcomes for increasing neural
299 sensitivities (Figure 2E). Probability modulations as tested in the current study are rather static and
300 depend on word probability in language, which has been learned over the course of one's life. It would be
301 interesting to also investigate whether these effects are dynamic by varying event probabilities within the
302 course of an experiment. If this manipulation also leads to similar phase-dependent categorization, phase
303 codes would not only be adjusted solely based on long-term hard-coded changes in excitability, but also
304 based on dynamically changing excitability levels that rely on top-down feedback¹⁷.

305 It has been proposed that speech comprehension involves first parsing speech into separate
306 temporal chunks using neural oscillations⁴³⁻⁴⁵. This segmentation is hypothetically done by aligning theta
307 band oscillations with syllables in an ongoing speech stream⁴⁴. In this way, one can parse and identify
308 individual syllables and use them for higher order linguistic operations⁴⁶. In our study, it is difficult to
309 separate segmentation or 'chunking' from any kind of process of interpretation. If a word is first
310 'segmented' or 'chunked' by theta oscillations, the information about the phase would be lost in

311 subsequent operations. However, in our study, it is exactly the phase of the theta oscillation that
312 determines how a word is interpreted. Note also that the reported effects are strikingly close to the 6.25
313 Hz frequency we found in our previous study¹¹. Therefore, our study shows that segmentation or chunking
314 through oscillations cannot be treated as a wholly separate process from word recognition, because
315 oscillations also provide a categorization mechanism alongside any potential segmentation or ‘chunking’
316 operation (see also ^{47,48|49}).

317 We have previously argued that during natural speech and language processing temporal
318 information can be used to infer information content, in other words, time can be a cue for content (also
319 see ^{17,50}). This time-content relation is governed by the observation that words that are more likely in the
320 current context are uttered with shorter inter-word-intervals^{17,51,52}. Combining this observation with
321 theories of oscillatory tracking results in more likely words being naturally presented at earlier, less
322 excitable phases, which we confirmed with our computational model¹⁷. This type of phase code can aid
323 speech comprehension: when information is ambiguous, the phase of an oscillation, and thereby the time
324 of word presentation, can be used to determine the percept. Notably, our current study does put some
325 limitations on this use of phase-dependent categorization. Our behavioral analysis shows that this phase-
326 dependent categorization works mostly when trait features across the word are all either frequent or non-
327 frequent. Yet, it is not clear how likelihood information that varies based on the level of linguistic analysis
328 interacts with timing in natural speech. To investigate this, one would need to show how top-down
329 feedback modulates phase-dependent categorization in sentence context as top-down feedback strongly
330 influences the sensitivity of individual word nodes as well as the timing of speech¹⁷.

331 In our computational model, we postulated a separate phonological and lexical level of analysis.
332 This is in line with neuroscientific results, which show that speech analysis is split up in separate analytical
333 levels across the cortex^{23,24,53}. Specifically, it is suggested that STG is sensitive to phonological content,
334 while MTG is sensitive to lexical access and word content. It is known that anatomical and functional
335 connectivity between these regions operate via early auditory cortex. There are direct connections
336 between early auditory cortex and STG and MTG, but also connections between STG and MTG⁵⁴⁻⁵⁶. Our
337 model currently relies primarily on the direct connections. An additional area strongly involved in the
338 language network is the IFG. We could not find any phase-dependent categorization in the IFG. This null
339 finding is not necessarily surprising as IFG has mostly been associated with higher-order language
340 processes that involve temporal integration across words and syntactic analysis^{27,28,57}. IFG might simply
341 be less sensitive to phonological or word-level likelihood differences. Alternatively, it is possible that we

342 did not find phase-dependent effects simply because there is too much temporal variability in the IFG
343 responses to speech. As we map phase of a cortical region directly to the onset of the presented sound,
344 any variability will reduce the accuracy of the phase estimation in relation with the behavioral outcome.
345 It can be expected that an area high up the processing hierarchy such as IFG has relatively high temporal
346 variability of the neural response to speech.

347 To conclude, we show that word categorization depends on the oscillatory phase of the neural
348 populations in regions where lexical and phonological information effects are typically observed. The
349 categorization bias is mediated by the trait or linguistic unit frequency that is putatively represented in
350 the region, perhaps through variation in sensitivity of diverse neural populations. We find a double
351 dissociation in which the phase in STG biases participants to the low or high frequent consonant percept,
352 while the phase in MTG biases participants to the low or high frequent word percept. These results
353 demonstrate that oscillations provide a temporal ordering of neural activation based on the excitability
354 of neural populations. Moreover, our study highlights the role of low frequency oscillations to organize
355 neural activation patterns along a gradient of event probability and provide an outlook for further
356 investigating the fundamental mechanisms that may be expressed via population rhythmic activity.

357 **Online Methods**

358 **Computational modelling**

359 We used a modified version of the Speech Tracking in a Model Constrained Oscillatory Network
360 (STiMCON) model¹⁷. In this model, a population of neural nodes is modulated by an oscillation and
361 individual nodes are additionally modulated based on their connectivity pattern with sensory input (see
362 formula (1). In the original model individual nodes are also modulated based on feedback). Each node is
363 governed by a non-linear inhibition function:

$$364 \quad \text{inhib}(Ta) = \begin{cases} -3 * \text{BaseInhib}, Ta < 20 \\ 3 * \text{BaseInhib}, 20 \leq Ta \leq 100 \\ \text{BaseInhib}, Ta > 100 \end{cases} \quad (2)$$

365 in which BaseInhib is a constant factor for the base inhibition level (set to -0.2, same as in ¹⁷). Initiation of
366 the inhibition function is governed by the activation threshold (by default set on 1, but varies with neural
367 sensitivity, see main text). First, this function creates suprathreshold activation after which nodes are
368 inhibited. The oscillatory function is as follows:

$$369 \quad \text{osc}(T) = Am * e^{2\pi i \omega T + i\varphi} \quad (3)$$

370 in which Am is the amplitude of the oscillator (set to 1), ω the frequency (set to 6.25Hz in accordance with
371 ¹¹), and φ the phase offset (variable). For the psychophysics experiment the phase offset is equalized with
372 the phase of the stimulus input. In the model, sensory input is directed to two different levels of analysis,
373 a phonetic level and a word level of analysis. Sensory input itself is modelled as a linear function ranging
374 between 0 to 1 in arbitrary units to the individual nodes lasting half a cycle. The maximum strength of the
375 sensory input depends on the morph level presented (see main text).

376

377 **Behavioral experiment**

378 *Participants.* In total 36 (28 female; age range = 18-40; mean age = 24.3) and 28 (20 female; age range =
379 20-59; mean age = 27.6) Dutch native speakers completed the session for the consonant and vowel
380 experiment respectively. All participants reported normal hearing and did not have any history of
381 language related disorders. Participants gave informed consent online. The study was approved by the
382 Ethics Board of the Social Sciences Faculty of Radboud University in Nijmegen. Participants received
383 monetary reimbursement for their participation.

384 *Materials.* Google text-to-speech was used to utter the Dutch word *daad* (IPA [international
385 phonetic alphabet]: /dat/; translation: *deed*) and *gaat* (IPA: /xat/; translation: *go*). We spliced the audio-
386 file to only contain the /da/ and /xa/ parts. We max-normalized these spliced audio fragments. In praat⁵⁸
387 we equalized the pitch contours of /da/ and /xa/ to lie in the middle of the original pitch contours of the
388 two sounds. For the consonant manipulation we morphed the two sounds together by taking a weighted
389 average of the two audio fragments in 11 spaced steps. Note that this step is different than in our original
390 study¹¹ in which we changed the formants directly, but it was necessary as we could not achieve a guttural
391 /x/ made in Dutch by using a formant change only. While for syllable perception this procedure is not a
392 problem (even though the sound is then closer to a /ga/ than a /xa/), for the chosen words a guttural /x/
393 is needed to understand the words. For the vowel manipulation, we subsequently changed the temporal
394 modulation (for all 11 morphs) of the original /a/ sound to be 0.75 of the original duration using PSOLA⁵⁹
395 (which can maintain pitch while changing temporal rate). Then we changed the spectral content of the
396 second formant in 11 steps from 1300-1700 Hz during the vowel utterance using the burgs LPC method⁶⁰.
397 This morph generates the phoneme /a/. We again max-normalized the output of these morphs. We
398 spliced from the original /dat/ sound fragment the /t/, shortened it to 0.9 of the original length (to
399 improve the sound audibility) and concatenated the /t/ at the end of all created morphs. The amplitude
400 of the last 0.2 seconds was linearly dampened. This whole procedure created a total of 11×11 morphs,
401 morphing between the four extreme sounds /xat/, /dat/, /xat/, and /dat/. Note that from all these morphs
402 we only used a total of 11×4–4 sounds, that is, the sounds at which either the consonant or vowel was at
403 its most extreme value. We choose these words as they are dissociable in vowel, consonant and word
404 frequency (see table 1).

405 For the psychophysics experiment, we presented a rhythmic sequence of broadband noise at 6.25
406 Hz before the word. Broadband noise consisted of 50 ms (with 5 ms linear amplitude ramp up and down)
407 between 1.1 and 3.1 kHz. Sequences lasted randomly 2, 3, or 4 seconds. Stimulus onset asynchrony (SOAs)
408 of the word relative to the final noise in the sequence was set to be between 0.1 to 0.42 in 12 equidistant
409 steps (covering two cycles of 6.25 Hz).

410 *Procedure.* All procedures were done online using Django web development, running under
411 Apache. In the first part of the experiment, we determined the most ambiguous stimulus as the morph
412 for which participants heard one of the two extremes of the morph 50% of the time. In both experiments,
413 this entails two morph spectra: for the consonant experiment the /dat-/xat/ and /dat-/xat/ spectra; for
414 the vowel experiment the /dat-/dat/ and /xat-/xat/ spectra. To do so, we presented all 22 morphed

415 sounds for the respective experiment and participants had to indicate what word they heard. A trial
416 consisted of a silent period of 0.5 seconds followed by the presentation of the audio fragment. 0.25
417 seconds after the sound, participants viewed the response options and could indicate via a button press
418 which sound they heard. Participants received two response options corresponding to the two extremes
419 of the spectrum to which the sound morph belonged to. In total, each sound was presented 12 times,
420 corresponding to 264 trials divided into two blocks. In total, the first part lasted about 8 minutes.
421 Immediately after this part for both spectra a psychometric logit function was fitted, and the most
422 ambiguous sound was determined.

423 The second, main, part of the experiment consisted of 14 blocks in which we presented the
424 rhythmic sequences with the final words. In total we presented 648 experimental trials: nine repetitions,
425 three sequence lengths, twelve SOAs, and two sound types. We added 5% of filler trials consisting of the
426 extreme sound types (at random sequence length and SOAs) resulting in a total of 680 trials. These trials
427 were added to test that participants were performing the task and not randomly pressing buttons and to
428 make sure that in some instances there was also a clear correct answer. Again, after the button press
429 there was an interval of 0.5 seconds. Throughout the experiment we adapted the ambiguous sound when
430 participants heard the same sound too often in a row. Specifically, if participants categorized the
431 ambiguous sound as the same word for 10 times in a row (for each spectra), we adjusted the ambiguous
432 sound one morph step away from the perceived word.

433 *Behavioral analysis.* For the first part of the experiment, we fitted a psychometric curve using the
434 `curve_fit` function in the `scipy` toolbox in python and extracted the most ambiguous stimulus for each
435 participant. We had difficulty to ensure that participants maintained an ambiguous percept either during
436 the psychometric determination or during the main experiment. We therefore had to exclude quite a few
437 participants from the analysis. This is likely due to the online procedure that needed to be done during
438 the COVID pandemic to which we had to rely on the audio of the participants at their home situation. As
439 the experimental results hinge on having an ambiguous percept, we needed to exclude participants who
440 could not maintain an ambiguous word perception throughout the experiment. During the first part of
441 the experiment, we could not fit an ambiguous sound that stood in between the 10th and 90th percentile
442 of the morph spectrum for 7 and 4 participants for consonant and vowel experiment, respectively. Also,
443 during the main experiment, 6 and 11 participants reported low perceived differences between the two
444 unambiguous words in the corresponding spectrum, respectively (under 20% difference between the
445 answers to the two unambiguous words; likely due to a failure to comply with the task or difficulties with

446 the task itself). An additional 5 and 1 participants reached morphs outside the 10th and 90th percentile
447 ranges for more than half the duration of the main experiment, respectively. This ended us with 18 and
448 12 participants for the consonant and vowel experiment, respectively. Trials in which morph estimations
449 were outside of the 10th and 90th percentile ranges were excluded from the analysis.

450 For those participants who met our criteria, we created a time course across the twelve SOAs
451 used. For each participant we subtracted the mean of the time course. Then we averaged all time-courses
452 and fitted a 6.25 Hz sinusoid to the data (with varying amplitude and phase) and extracted the explained
453 variance. We generated a null distribution by randomly permuting ($n = 10,000$) the average time course
454 and refitting the sinusoid. One-sided p-values were extracted by comparing the proportion of the
455 observed explained variance with the explained variance of the null distribution.

456

457 **MEG experiment**

458 *Participants.* 23 Dutch native speakers (13 females; age range: 18-59; mean age = 34.3; one author
459 participated as well) participated in the study. 22 were right-handed (one reported no preference in
460 hand). All reported normal hearing, had normal or corrected-to-normal vision, and did not have any
461 history of dyslexia or other language-related disorders. Participants performed a screening for their
462 eligibility in the MEG and MRI and gave written informed consent. The study was approved by the ethical
463 Commission for human research Arnhem/Nijmegen (project number CMO2014/288). All participants
464 were reimbursed for their participation. One participant was excluded for not maintaining an ambiguous
465 percept throughout the experiment.

466 *Procedure.* Just as the behavioral experiment, the MEG experiment consisted of two parts. In the
467 first part, we repeated the presentation of the psychometric function to determine the most ambiguous
468 sounds. After the sound was finished the response options were immediately shown. The next sound was
469 presented at random interval between 0.5-1.5 seconds after the response. In the MEG experiment, we
470 only used the consonant morphs. During the main part of the experiment, we presented 50% of the time
471 ambiguous ($n = 160$ per spectrum) and 50% of the time non-ambiguous sounds ($n = 80$ for each extreme
472 per spectrum). All trial types were presented in pseudo-random order. After sound presentation, there
473 was a 1 second interval before the sound options were shown. The next sound was presented at a random
474 interval between 2 and 4 seconds after the response. After 80 sounds participants had a break. The
475 response options of participants (pressing left or right for /d/ vs /x/, respectively) were switched halfway

476 through the experiment to ensure that the effects were not due to motor plans. Half of the participants
477 started with /d/ as the left option and the other half with /x/ as the left option. During the experiment,
478 the morph was changed if the participant reported the same percept for the ambiguous sound within one
479 spectrum for 10 consecutive answers. At the end of the experiment, we collected an auditory localizer
480 (data not analyzed here) and a scalp digitization using the Polhemus Fastrak digitizer. All stimulus
481 presentation was programmed in Psychtoolbox⁶¹ and run in the linux environment.

482 *MEG pre-processing.* Surface-based source models from the MRI were made using grid points that
483 were defined on the cortical sheet of the automatic segmentation of freesurfer6.0 [50] in combination
484 with pre-processing tools from the HCP workbench1.3.2 [51] to down-sample the mesh to 4k vertices per
485 hemisphere. The MRI was co-registered to the MEG by using the previously defined fiducials as well as an
486 automatic alignment of the MRI to the Polhemus headshape using the Fieldtrip20211102 software [52].
487 Head models were based on the SPM segmentation incorporated in Fieldtrip. Regions of interest (ROIs)
488 were the superior temporal gyrus (STG), the middle temporal gyrus (MTG) and the inferior frontal cortex
489 (IFG). Parcellations were based on the Freesurfer parcellations.

490 Preprocessing of MEG involved epoching the data both between -2 and 0 seconds and -1 to +1
491 seconds relative to sound onset. This separate epoching was necessary to ensure that for the pre-stimulus
492 analyses no data from the post-stimulus interval could leak into the pre-stimulus interval due to filtering
493 during preprocessing. Data was then padded for 0.5 seconds at the end of the epoch with the last value
494 of the epoch. Data was low-passed at 100 Hz and DFT notch filters were applied at 50, 100 and 150 Hz.
495 Then, data was down-sampled to 300 Hz and the padded interval was removed again. ICA was performed
496 to remove heartbeat-related signals and eye blinks and movements. On average 4.3 components (range:
497 3-6 components) were removed from the analysis. After that, trials with excessive noise were removed
498 via visual inspection with an average of 12.7 removed trials (range: 4-22 trials). We calculated a common
499 spatial filter using lcmv filter based on the post-stimulus data with a lambda of 5%. Many spatial filters
500 have a center of the head bias, resulting in stronger activity in the center compared to the cortical
501 surface²⁹. This bias is often counteracted by having a clear baseline period for each trial to which the data
502 is referenced to. However, for our analysis no clear baseline period can be defined as we were interested
503 in the pre-stimulus period. Therefore, to counteract the center of head bias we used an array-gain
504 beamformer which normalized the spatial filter²⁹. This filter was applied to all single trial estimates of the
505 pre-stimulus data. To extract a single time course representative of our ROIs we extracted the first PCA
506 for each ROI.

507 *MEG analysis:* We performed a time-frequency analysis on all source trials using a wavelet
508 approach extracting frequencies from 1 to 15 Hz in steps of 0.5 Hz at widths matching 700 milliseconds
509 for the timepoints -0.5 up to 0 seconds in steps of 0.05 seconds extracting the phase of the complex
510 Fourier spectra. Data corresponding to the ambiguous sounds were then split according to the response
511 of the participants based on two different contrast binning:

- 512 1) Consonant frequency binning: responses where the ambiguous word was interpreted as a word
513 with a low frequency consonant (/xat/ and /xat/) versus a word with a high frequency consonant
514 (/dat/ and /dat/).
- 515 2) Word frequency binning: responses where the ambiguous word was interpreted as a low
516 frequency word (/dat/ and /xat/) versus a high frequency word (/dat/ and /xat/).

517 For these two contrasts, low frequency traits were labeled with a zero and high frequency traits
518 with a one. A circular-linear correlation was performed between the pre-stimulus phase and the response
519 of the participant for all time and frequency points. This analysis produced a correlation value and a p-
520 value for each participant. As the exact correlation value is influenced by the number of trials included in
521 the correlation, we performed statistics using the inverse of the normative cumulative distribution based
522 on the p-value of the correlation of the individual participants (also see⁶²). Group statistics were
523 performed by statistically testing these z-values against zero using a one-sample t-test. To control for
524 multiple comparisons across all time-frequency points we performed cluster-based permutation tests³¹.

525 To further inspect the effect of the two contrasts we split the data based on all four possible
526 response options. Then, we extracted for each participant and for each of the possible perceived words
527 the average phase at which participants reported perceiving that specific word. We calculated the phase
528 difference between the average phase at which participants perceived the word /dat/ and the other three
529 options. The logic of this analysis was as follows: /dat/ has high frequency features for all investigated
530 feature dimensions. Words that also have a high frequency content should therefore show a phase
531 difference of zero with /dat/, but words that have a low frequency content should show a phase difference
532 of π with /dat/. We statistically tested whether the phases were non-uniform around the expected phase
533 using the v-test statistic^{63,64}. This test will show a significant effect only when the data is both non-uniform
534 and the phase is around the expected phase.

535 For the power analysis, we used the same wavelet approach as for the time-frequency plots
536 (Figure 4C) and converted the data in z-scores across the whole time-frequency window. For the power

537 spectra (Figure 4B), we cut the data from -1 to 0, padded the data out to 5 seconds and extracted the
538 power using Hanning tapers.

539 Acknowledgements

540 AEM was supported by the Lise Meitner Research Group “Language and Computation in Neural Systems”,
541 by NWO Vidi grant 016.Vidi.188.029, and by the Language in Interaction Consortium which is funded by
542 NWO Gravitation Grant 024.001.006.

543

544

545 References

- 546 1 Haegens, S., Nacher, V., Luna, R., Romo, R. & Jensen, O. alpha-Oscillations in the monkey
547 sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal
548 spiking. *Proc Natl Acad Sci U S A* **108**, 19377-19382 (2011).
549 [https://doi.org:10.1073/pnas.1117190108](https://doi.org/10.1073/pnas.1117190108)
- 550 2 Buzsaki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science* **304**, 1926-1929
551 (2004). [https://doi.org:10.1126/science.1099745](https://doi.org/10.1126/science.1099745)
- 552 3 Schroeder, C. E. & Lakatos, P. Low-frequency neuronal oscillations as instruments of sensory
553 selection. *Trends Neurosci.* **32**, 9-18 (2009). [https://doi.org:10.1016/j.tins.2008.09.012](https://doi.org/10.1016/j.tins.2008.09.012)
- 554 4 Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M. & Ro, T. To see or not to see: prestimulus
555 alpha phase predicts visual awareness. *J. Neurosci.* **29**, 2725-2732 (2009).
556 [https://doi.org:10.1523/JNEUROSCI.3963-08.2009](https://doi.org/10.1523/JNEUROSCI.3963-08.2009)
- 557 5 Dugué, L., Marque, P. & VanRullen, R. The phase of ongoing oscillations mediates the causal
558 relation between brain excitation and visual perception. *J. Neurosci.* **31**, 11889-11893 (2011).
- 559 6 Henry, M. J., Herrmann, B. & Obleser, J. Entrained neural oscillations in multiple frequency bands
560 comodulate behavior. *Proc Natl Acad Sci U S A* **111**, 14935-14940 (2014).
561 [https://doi.org:10.1073/pnas.1408741111](https://doi.org/10.1073/pnas.1408741111)
- 562 7 Busch, N. A., Dubois, J. & VanRullen, R. The phase of ongoing EEG oscillations predicts visual
563 perception. *J. Neurosci.* **29**, 7869-7876 (2009). [https://doi.org:10.1523/JNEUROSCI.0113-09.2009](https://doi.org/10.1523/JNEUROSCI.0113-09.2009)
- 564 8 Ten Oever, S., De Weerd, P. & Sack, A. T. Phase-dependent amplification of working memory
565 content and performance. *Nat Commun* **11**, 1832 (2020). [https://doi.org:10.1038/s41467-020-](https://doi.org/10.1038/s41467-020-15629-7)
566 [15629-7](https://doi.org/10.1038/s41467-020-15629-7)
- 567 9 Strauss, A., Henry, M. J., Scharinger, M. & Obleser, J. Alpha phase determines successful lexical
568 decision in noise. *J. Neurosci.* **35**, 3256-3262 (2015). [https://doi.org:10.1523/JNEUROSCI.3357-](https://doi.org/10.1523/JNEUROSCI.3357-14.2015)
569 [14.2015](https://doi.org/10.1523/JNEUROSCI.3357-14.2015)
- 570 10 Kayser, S. J., McNair, S. W. & Kayser, C. Prestimulus influences on auditory perception from
571 sensory representations and decision processes. *Proc Natl Acad Sci U S A* **113**, 4842-4847 (2016).
572 [https://doi.org:10.1073/pnas.1524087113](https://doi.org/10.1073/pnas.1524087113)
- 573 11 ten Oever, S. & Sack, A. T. Oscillatory phase shapes syllable perception. *Proc Natl Acad Sci U S A*
574 **112**, 15833-15837 (2015). [https://doi.org:10.1073/pnas.1517519112](https://doi.org/10.1073/pnas.1517519112)
- 575 12 Theze, R., Giraud, A. L. & Megevand, P. The phase of cortical oscillations determines the
576 perceptual fate of visual cues in naturalistic audiovisual speech. *Sci Adv* **6**, eabc6348 (2020).
577 [https://doi.org:10.1126/sciadv.abc6348](https://doi.org/10.1126/sciadv.abc6348)
- 578 13 Fries, P. A mechanism for cognitive dynamics: neuronal communication through neuronal
579 coherence. *Trends Cogn. Sci.* **9**, 474-480 (2005). [https://doi.org:10.1016/j.tics.2005.08.011](https://doi.org/10.1016/j.tics.2005.08.011)

- 580 14 Mehta, M. R., Lee, A. K. & Wilson, M. A. Role of experience and oscillations in transforming a rate
581 code into a temporal code. *Nature* **417**, 741-746 (2002). <https://doi.org/DOI>
582 [10.1038/nature00807](https://doi.org/10.1038/nature00807)
- 583 15 McGann, J. P. Associative learning and sensory neuroplasticity: how does it happen and what is it
584 good for? *Learn. Memory* **22**, 567-576 (2015).
- 585 16 Summerfield, C. & de Lange, F. P. Expectation in perceptual decision making: neural and
586 computational mechanisms. *Nat. Rev. Neurosci.* **15**, 745-756 (2014).
587 <https://doi.org/10.1038/nrn3838>
- 588 17 Ten Oever, S. & Martin, A. E. An oscillating computational model can track pseudo-rhythmic
589 speech by using linguistic predictions. *Elife* **10**, e68066 (2021).
590 <https://doi.org/10.7554/eLife.68066>
- 591 18 Ten Oever, S., Sack, A. T., Wheat, K. L., Bien, N. & van Atteveldt, N. Audio-visual onset differences
592 are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Front Psychol*
593 **4**, 331 (2013). <https://doi.org/10.3389/fpsyg.2013.00331>
- 594 19 Zuidema, W. A syllable frequency list for Dutch. (2010).
- 595 20 Lisman, J. E. & Jensen, O. The Theta-Gamma Neural Code. *Neuron* **77**, 1002-1016 (2013).
596 <https://doi.org/10.1016/j.neuron.2013.03.007>
- 597 21 Panzeri, S., Macke, J. H., Gross, J. & Kayser, C. Neural population coding: combining insights from
598 microscopic and mass signals. *Trends Cogn. Sci.* **19**, 162-172 (2015).
599 <https://doi.org/10.1016/j.tics.2015.01.002>
- 600 22 Haegens, S. & Zion Golumbic, E. Rhythmic facilitation of sensory processing: A critical review.
601 *Neurosci. Biobehav. Rev.* **86**, 150-165 (2018). <https://doi.org/10.1016/j.neubiorev.2017.12.002>
- 602 23 Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**,
603 393-402 (2007). <https://doi.org/10.1038/nrn2113>
- 604 24 Gow, D. W., Jr. The cortical organization of lexical knowledge: a dual lexicon model of spoken
605 language processing. *Brain Lang* **121**, 273-288 (2012).
606 <https://doi.org/10.1016/j.bandl.2012.03.005>
- 607 25 Brennan, J. R. *Language and the Brain: A Slim Guide to Neurolinguistics*. (Oxford University Press,
608 2022).
- 609 26 Lewis, G. & Poeppel, D. The role of visual representations during the lexical access of spoken
610 words. *Brain Lang* **134**, 1-10 (2014). <https://doi.org/10.1016/j.bandl.2014.03.008>
- 611 27 Dehaene, S., Meyniel, F., Wacongne, C., Wang, L. & Pallier, C. The Neural Representation of
612 Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron* **88**, 2-
613 19 (2015). <https://doi.org/10.1016/j.neuron.2015.09.019>
- 614 28 Park, H., Ince, R. A., Schyns, P. G., Thut, G. & Gross, J. Frontal top-down signals increase coupling
615 of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* **25**,
616 1649-1653 (2015). <https://doi.org/10.1016/j.cub.2015.04.049>
- 617 29 Westner, B. U. *et al.* A unified view on beamformers for M/EEG source reconstruction.
618 *NeuroImage* **246**, 118789 (2022). <https://doi.org/10.1016/j.neuroimage.2021.118789>
- 619 30 Sekihara, K. & Nagarajan, S. S. *Adaptive spatial filters for electromagnetic brain imaging*. (Springer
620 Science & Business Media, 2008).
- 621 31 Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci.*
622 *Methods* **164**, 177-190 (2007). <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- 623 32 Landau, A. N. & Fries, P. Attention samples stimuli rhythmically. *Curr. Biol.* **22**, 1000-1004 (2012).
624 <https://doi.org/10.1016/j.cub.2012.03.054>
- 625 33 Fiebelkorn, I. C., Saalman, Y. B. & Kastner, S. Rhythmic Sampling within and between Objects
626 despite Sustained Attention at a Cued Location. *Curr. Biol.* **23**, 2553-2558 (2013).
627 <https://doi.org/10.1016/j.cub.2013.10.063>

- 628 34 Stefanics, G. *et al.* Phase Entrainment of Human Delta Oscillations Can Mediate the Effects of
629 Expectation on Reaction Speed. *J. Neurosci.* **30**, 13578-13585 (2010).
630 <https://doi.org/10.1523/Jneurosci.0703-10.2010>
- 631 35 Schaefer, A. T., Angelo, K., Spors, H. & Margrie, T. W. Neuronal oscillations enhance stimulus
632 discrimination by ensuring action potential precision. *PLoS Biol.* **4**, e163 (2006).
633 <https://doi.org/10.1371/journal.pbio.0040163>
- 634 36 O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta
635 rhythm. *Hippocampus* **3**, 317-330 (1993).
- 636 37 Maurer, A. P. & McNaughton, B. L. Network and intrinsic cellular mechanisms underlying theta
637 phase precession of hippocampal neurons. *Trends Neurosci.* **30**, 325-333 (2007).
638 <https://doi.org/10.1016/j.tins.2007.05.002>
- 639 38 Martin, A. E. & Dumas, L. A. Predicate learning in neural systems: using oscillations to discover
640 latent structure. *Current Opinion in Behavioral Sciences* **29**, 77-83 (2019).
- 641 39 Dumas, L. A., Hummel, J. E. & Sandhofer, C. M. A theory of the discovery and predication of
642 relational concepts. *Psychological review* **115**, 1 (2008).
- 643 40 Martin, A. E. & Dumas, L. A. Tensors and compositionality in neural systems. *Philosophical
644 Transactions of the Royal Society B* **375**, 20190306 (2020).
- 645 41 Jensen, O. & Lisman, J. E. Hippocampal CA3 region predicts memory sequences: accounting for
646 the phase precession of place cells. *Learn Mem* **3**, 279-287 (1996). [https://doi.org/10.1101/lm.3.2-
647 3.279](https://doi.org/10.1101/lm.3.2-3.279)
- 648 42 Heusser, A. C., Poeppel, D., Ezzyat, Y. & Davachi, L. Episodic sequence memory is supported by a
649 theta-gamma phase code. *Nat. Neurosci.* **19**, 1374+ (2016). <https://doi.org/10.1038/nn.4374>
- 650 43 Giraud, A. L. & Poeppel, D. Cortical oscillations and speech processing: emerging computational
651 principles and operations. *Nat. Neurosci.* **15**, 511-517 (2012). <https://doi.org/10.1038/nn.3063>
- 652 44 Ghitza, O. On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech
653 with a manipulated modulation spectrum. *Front Psychol* **3**, 238 (2012).
654 <https://doi.org/10.3389/fpsyg.2012.00238>
- 655 45 Meyer, L. The neural oscillations of speech processing and language comprehension: state of the
656 art and emerging mechanisms. *Eur. J. Neurosci.* **48**, 2609-2621 (2018).
657 <https://doi.org/10.1111/ejn.13748>
- 658 46 Martin, A. E. & Dumas, L. A. A mechanism for the cortical computation of hierarchical linguistic
659 structure. *PLoS Biol.* **15**, e2000663 (2017).
- 660 47 Bai, F., Meyer, A. S. & Martin, A. E. Neural dynamics differentially encode phrases and sentences
661 during spoken language comprehension. *PLoS Biol.* **20**, e3001713 (2022).
- 662 48 Brennan, J. R. & Martin, A. E. Phase synchronization varies systematically with linguistic structure
663 composition. *Philosophical Transactions of the Royal Society B* **375**, 20190305 (2020).
- 664 49 Luo, H. & Poeppel, D. Phase patterns of neuronal responses reliably discriminate speech in human
665 auditory cortex. *Neuron* **54**, 1001-1010 (2007).
- 666 50 Martin, A. E. A compositional neural architecture for language. *J. Cognit. Neurosci.*, 1-20 (2020).
- 667 51 Aylett, M. P. & Turk, A. The smooth signal redundancy hypothesis: A functional expla-
668 nation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech.
669 *Language and Speech* **47**, 31-56 (2004).
- 670 52 Seyfarth, S. Word informativity influences acoustic duration: effects of contextual predictability
671 on lexical representation. *Cognition* **133**, 140-155 (2014).
672 <https://doi.org/10.1016/j.cognition.2014.06.013>

- 673 53 Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W. M. & Hale, J. T. Abstract linguistic
674 structure correlates with temporal activity during naturalistic comprehension. *Brain Lang* **157-**
675 **158**, 81-94 (2016). <https://doi.org/10.1016/j.bandl.2016.04.008>
- 676 54 Xu, J. *et al.* Tractography-based Parcellation of the Human Middle Temporal Gyrus. *Sci Rep* **5**,
677 18883 (2015). <https://doi.org/10.1038/srep18883>
- 678 55 Glasser, M. F. & Rilling, J. K. DTI tractography of the human brain's language pathways. *Cereb.*
679 *Cortex* **18**, 2471-2482 (2008). <https://doi.org/10.1093/cercor/bhn011>
- 680 56 Fernandez, L., Velasquez, C., Garcia Porrero, J. A., de Lucas, E. M. & Martino, J. Heschl's gyrus fiber
681 intersection area: a new insight on the connectivity of the auditory-language hub. *Neurosurg*
682 *Focus* **48**, E7 (2020). <https://doi.org/10.3171/2019.11.FOCUS19778>
- 683 57 Ten Oever, S., Carta, S., Kaufeld, G. & Martin, A. E. Neural tracking of phrases in spoken language
684 comprehension is automatic and task-dependent. *Elife* **11**, e77468 (2022).
685 <https://doi.org/10.7554/eLife.77468>
- 686 58 Praat: a system for doing phonetics by computer v. 5.3.56 (2013).
- 687 59 Moulines, E. & Charpentier, F. Pitch-synchronous waveform processing techniques for text-to-
688 speech synthesis using diphones. *Speech communication* **9**, 453-467 (1990).
- 689 60 Andersen, N. On the calculation of filter coefficients for maximum entropy spectral analysis.
690 *Geophysics* **39**, 69-72 (1974).
- 691 61 Brainard, D. H. The Psychophysics Toolbox. *Spat Vis* **10**, 433-436 (1997).
- 692 62 Cohen, M. X. *Analyzing neural time series data: theory and practice*. (MIT press, 2014).
- 693 63 Zar, J. H. *Biostatistical Analysis*. 4 edn, (Prentice Hall, 1998).
- 694 64 Fisher, N. I. *Statistical analysis of circular data*. (Cambridge University Press, 1995).

695