1 **Title:** The categorical neural organization of speech aids its perception in noise

2

3

4

5

6 **Author(s):** Gavin M. **Bidelman**[1,2,3,†], Lauren C. **Bush**[2,a], & Alex M. **Boudreaux**[2,a]

7

8

9 [1]Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA

10 [2]School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA

11 [3]University of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis,

12 TN, USA

13

14

15 [a] These authors contributed equally to this work

16

17

18

19

20

21

22

23

24

25

26

27 [†]**Address for editorial correspondence:**
28 Gavin M. Bidelman, PhD
29 School of Communication Sciences & Disorders
30 University of Memphis
31 4055 North Park Loop
32 Memphis, TN, 38152
33 TEL: (901) 678-5826
34 FAX: (901) 525-1282
35 EMAIL: gmbdlman@memphis.edu

36 **ABSTRACT**

37 We investigated whether the categorical perception (CP) of speech might also provide a mechanism that

38 aids its perception in noise. We varied signal-to-noise ratio (SNR) [clear, 0 dB, -5 dB] while listeners

39 classified an acoustic-phonetic continuum (/u/ to /a/). Noise-related changes in behavioral categorization

40 were only observed at the lowest SNR. Event-related brain potentials (ERPs) differentiated phonetic vs.

41 non-phonetic (category ambiguous) speech by the P2 wave (~180-320 ms). Paralleling behavior, neural

42 responses to speech with clear phonetic status (i.e., continuum endpoints) were largely invariant to noise,

43 whereas responses to ambiguous tokens declined with decreasing SNR. Results demonstrate that phonetic

44 speech representations are more resistant to degradation than corresponding acoustic representations.

45 Findings suggest the mere process of binning speech sounds into categories provides a robust mechanism

46 to aid perception at the "cocktail party" by fortifying abstract categories from the acoustic signal and

47 making the speech code more resistant to external interferences.

48

49

50

51 **Keywords:** Auditory event-related potentials (ERPs); categorical perception; speech-in-noise (SIN)

52 perception

53

54

55

56

57

58

59

## 1. INTRODUCTION

A basic tenet of perceptual organization is that sensory phenomena are subject to invariance: similar features are mapped to common identities (equivalence classes) by assigning similar objects to the same membership (Goldstone & Hendrickson, 2010), a process known as categorical perception (CP). In the context of speech, CP is demonstrated when gradually morphed sounds along an equidistant acoustic continuum are heard as only a few discrete classes (Bidelman et al., 2013; Harnad, 1987; Liberman et al., 1967; Pisoni, 1973; Pisoni & Luce, 1987). Equal physical steps along a signal dimension do not produce equivalent changes in percept (Holt & Lotto, 2006). Rather, listeners treat sounds within a given category as perceptually similar despite their otherwise dissimilar acoustics. Skilled categorization is particularly important for spoken and written language, as evidenced by its role in reading acquisition (Mody et al., 1997; Werker & Tees, 1987), sound-to-meaning learning (Myers & Swan, 2012; Reetzke et al., 2018), and putative deficits in language-based learning disorders (e.g., specific language impariment, dyslexia; Calcus et al., 2016; Noordenbos & Serniclaes, 2015; Werker & Tees, 1987). To arrive at categorical decisions, acoustic cues are presumably weighted and compared against internalized "templates" in the brain, built through repetitive exposure to one's native language (Bidelman & Lee, 2015; Guenther et al., 2004; Iverson et al., 2003a; Kuhl, 1991).

Beyond providing observers a smaller, more manageable perceptual space, CP might also aid degraded speech perception if phonetic categories are somehow more resistant to noise (Gifford et al., 2014; Helie, 2017). Indeed, categories (a higher-level code) are thought to be more robust to noise degradations than physical surface features of a signal (lower-level sensory code) (Bidelman et al., under review; Helie, 2017). A theoretical example of how categorical processing might aid the perception of degraded speech is illustrated in Figure 1.

[Insert Figure 1 near here]

Consider the neural representation of speech as a multidimensional feature space. Populations of auditory cortical neurons code different dimensions of the acoustic input. Categorical coding could be reflected as an increase (or conversely, decrease) in local firing rate for stimuli that are perceptually similar despite their otherwise dissimilar acoustics ("A" and "B") (e.g., Guenther & Gjaja, 1996; Guenther et al., 2004; Recanzone et al., 1993). Although noise interference would blur physical acoustic details and create a noisier cortical map, categories would be partially spared—indicated by the remaining "peakedness" in the neural space. Thus, both the construction of perceptual objects and natural discrete binning process of CP might enable category members to "pop out" among a noisy feature space (e.g., Nothdurft, 1991; Perez-Gay et al., 2018). Consequently, the mere process of grouping speech sounds into categories might aid comprehension of speech-in-noise (SIN)—assuming those representations are not

93 too severely compromised and remain distinguishable from noise itself. This theoretical framework

94 provides the basis for the current empirical study.

95   Building on our recent efforts to decipher the neurobiology of "cocktail party" listening and

96 understand the physiological mechanisms supporting robust speech perception (for review, see Bidelman,

97 2017), this study aimed to test whether speech sounds carrying strong phonetic categories are more

98 resilient to the deleterious effects of noise than categorically ambiguous speech sounds. When category-

99 relevant dimensions are less distinct and perceptual boundaries are particularly noisy, additional

100 mechanisms for enhancing separation must be engaged (Livingston et al., 1998). We hypothesized the

101 phonetic groupings inherent to speech may be one such mechanism. Because phonetic categories reflect a

102 more abstract, higher-level representation of speech (i.e., acoustic + phonetic code), we reasoned they

103 would be more robust to noise than physical features of speech that do not engage phonetic-level

104 processing (i.e., acoustic code) (cf. Bidelman et al., under review; Helie, 2017). To test this possibility,

105 we recorded high-density event-related potentials (ERPs) while listeners categorized speech continua in

106 different levels of acoustic noise. The critical comparison was between responses to stimuli at the

107 endpoints vs. midpoint of the acoustic-phonetic continuum. We predicted that if the categorization

108 process aids figure-ground perception, speech tokens having a clear phonetic identity (continuum

109 endpoints) would elicit lesser noise-related change in the ERPs than non-phonetic tokens (continuum

110 midpoint), which have a bistable (ambiguous) percept and lack a clear phonetic identity.

## 2. MATERIALS & METHODS

### 2.1 Participants

113   Fifteen young adults (3 male, 12 females; age: $M$ = 24.3, $SD$ = 1.7 years) were recruited from the

114 University of Memphis student body. All exhibited normal hearing sensitivity confirmed via a threshold

115 screening (i.e., < 20 dB HL, audiometric frequencies 250 - 8000 Hz). Each participant was strongly right-

116 handed (87.0± 18.2% laterality index; Oldfield, 1971) and had obtained a collegiate level of education

117 (17.8±1.9 years). Musical training is known to modulate categorical processing and SIN listening abilities

118 (Bidelman & Alain, 2015b; Bidelman & Krishnan, 2010; Bidelman et al., 2014b; Parbery-Clark et al.,

119 2009; Yoo & Bidelman, under review). Consequently, we required that all participants had minimal

120 music training throughout their lifetime (mean years of training: 1.3±1.8 years). All were paid for their

121 time and gave informed consent in compliance with the Declaration of Helsinki and a protocol approved

122 by the Institutional Review Board at the University of Memphis.

### 2.2 Speech continuum and behavioral task

124   We used a synthetic five-step vowel continuum to assess the neural correlates of CP (Bidelman &

125 Alain, 2015b; Bidelman & Walker, 2017; Bidelman et al., 2014b). Each token of the continuum was

126 separated by equidistant steps acoustically based on first formant frequency (F1), but was perceived

127 categorically from /u/ to /a/. Tokens were 100 ms, including 10 ms of rise/fall time to reduce spectral

128 splatter in the stimuli. Each contained identical voice fundamental (F0), second (F2), and third formant

129 (F3) frequencies (F0: 150, F2: 1090, and F3: 2350 Hz). The F1 was parameterized over five equal steps

130 between 430 and 730 Hz such that the resultant stimulus set spanned a perceptual phonetic continuum

131 from /u/ to /a/ (Bidelman et al., 2013). Speech stimuli were delivered binaurally at 75 dB SPL through

132 shielded insert earphones (ER-2; Etymotic Research) coupled to a TDT RP2 processor (Tucker Davis

133 Technologies).

134 This same speech continuum was presented in one of three noise blocks varying in signal-to-noise

135 ratio (SNR): clear, 0 dB SNR, -5 dB SNR (Fig. 2). These noise levels were selected based on extensive

136 pilot testing which confirmed they differentially hindered speech perception. The masker was a speech-

137 shaped noise based on the long-term power spectrum of the vowel set. Noise was presented continuously

138 so it was not time-locked to the stimulus presentation, providing a constant backdrop of acoustic

139 interference during the categorization task (e.g., Alain et al., 2012; Bidelman et al., 2018; Bidelman &

140 Howell, 2016). SNR was manipulated by changing the level of the masker to ensure SNR was inversely

141 correlated with overall sound level (Binder et al., 2004). Noise block order was randomized within and

142 between participants.

143            [Insert Figure 2 near here]

144 The task was otherwise identical to our previous neuroimaging studies on CP (e.g., Bidelman &

145 Alain, 2015b; Bidelman et al., 2013; Bidelman & Walker, 2017). During EEG recording, listeners heard

146 150 trials of each individual speech token (per noise block). On each trial, they were asked to label the

147 sound with a binary response ("u" or "a") as quickly and accurately as possible. Following the listener's

148 behavioral response, the interstimulus interval (ISI) was jittered randomly between 800 and 1000 ms (20

149 ms steps, uniform distribution) to avoid rhythmic entrainment of the EEG and the anticipation of

150 subsequent stimuli.

151 **2.3 EEG recording and preprocessing**

152 EEGs were recorded from 64 sintered Ag/AgCl electrodes at standard 10-10 scalp locations

153 (Oostenveld & Praamstra, 2001). Continuous data were digitized using a sampling rate of 500 Hz

154 (SynAmps RT amplifiers; Compumedics Neuroscan) and an online passband of DC-200 Hz. Electrodes

155 placed on the outer canthi of the eyes and the superior and inferior orbit monitored ocular movements.

156 Contact impedances were maintained < 10 kΩ during data collection. During acquisition, electrodes were

157 referenced to an additional sensor placed ~ 1 cm posterior to the Cz channel.

158    EEG pre-processing was performed in BESA® Research (v7) (BESA, GmbH). Ocular artifacts

159    (saccades and blinks) were first corrected in the continuous EEG using a principal component analysis

160    (PCA) (Picton et al., 2000). Cleaned EEGs were then filtered (1-30 Hz), epoched (-200-800 ms), baseline

161    corrected to the pre-stimulus interval, and averaged in the time domain resulting in 15 ERP waveforms

162    per participant (5 tokens * 3 noise conditions). For analysis, data were re-referenced using BESA's

163    reference-free virtual montage. This montage computes a spherical spline-interpolated voltage (Perrin et

164    al., 1989) for each channel relative to the mean voltage over 642 equidistant locations covering the entire

165    sphere of the head. This montage is akin to common average referencing but results in a closer

166    approximation to true reference free waveforms (Scherg et al., 2002). However, reported results were

167    similar using a common average reference (data not shown).

168    ERP quantification focused on the latency range of the P2 wave as previous studies have shown

169    the neural correlates of CP emerge around the timeframe of this component (Bidelman & Alain, 2015b;

170    Bidelman & Lee, 2015; Bidelman et al., 2013; Bidelman & Walker, 2017). Guided by visual inspection of

171    grand averaged data, we measured the amplitude of the evoked potentials as the positive-going deflection

172    between 180-320 ms. This window covered both the P2 and following P3b-like deflections (see Fig. 4).

173    To evaluate whether ERPs showed categorical coding, we averaged response amplitudes to prototypical

174    tokens at the endpoints of the continuum and compared this combination to the ambiguous token at its

175    midpoint (e.g., Bidelman, 2015; Bidelman & Walker, 2017; Liebenthal et al., 2010). This contrast (i.e.,

176    mean[Tk1, Tk5] vs. Tk3) allowed us to assess the degree to which noise affected the neural encoding of

177    phonetic categories (Tk1/5) vs. non-phonetic (Tk3) speech sounds and thus affected categorical

178    processing. The rationale for this analysis is that it effectively minimizes stimulus-related differences in

179    the ERPs, thereby isolating categorical/perceptual processing. For example, Tk1 and Tk5 are expected to

180    produce distinct ERPs due to exogenous processing alone. However, comparing the average of these

181    responses (i.e., mean[Tk1, Tk5]) to that of Tk3 allows us to better isolate ERP modulations related to

182    categorical coding (Bidelman & Walker, 2017).

183    Averaging endpoint responses doubles the number of trials for the prototypes relative to the

184    ambiguous condition, which could mean differences were attributable to SNR of the ERPs rather than CP

185    effects, *per se* (Hu et al., 2010). To rule out this possibility, we measured the SNR of the ERPs as

186    $10\log(RMS_{ERP}/RMS_{baseline})$ (Bidelman, 2018) where $RMS_{ERP}$ and $RMS_{baseline}$ were the RMS amplitudes of

187    the ERP (signal) portion of the epoch window (0-800 ms) and pre-response baseline period (-200-0ms

188    ms), respectively. Critically, SNR of the ERPs did not differ across conditions [$F_{5,70}$ =0.56, $p$=0.73],

189    indicating that neural activity was not inherently noisier for a given token type or acoustic noise level.

190    **2.4 Behavioral data analysis**

191    Identification scores were fit with a sigmoid function $P = 1/[1+e^{-\beta_1(x-\beta_0)}]$, where $P$ is the

192    proportion of trials identified as a given vowel, $x$ is the step number along the stimulus continuum, and $\beta_0$

193    and $\beta_1$ the location and slope of the logistic fit estimated using nonlinear least-squares regression.

194    Comparing parameters between SNR conditions revealed possible differences in the location and

195    "steepness" (i.e., rate of change) of the categorical boundary as a function of noise degradation. Larger $\beta_1$

196    values reflect steeper psychometric functions and thus stronger categorical perception.

197    Behavioral speech labeling speeds (i.e., reaction times [RTs]) were computed as listeners' median

198    response latency across trials for a given condition. RTs outside 250-2500 ms were deemed outliers (e.g.,

199    fast guesses, lapses of attention) and were excluded from the analysis (Bidelman et al., 2013; Bidelman &

200    Walker, 2017).

**2.5 Statistical analysis**

202    Unless otherwise noted, dependent measures were analyzed using a one-way, mixed model

203    ANOVA (subject=random factor) with fixed effects of SNR (3 levels: clear, 0 dB, -5dB) and token [5

204    levels: Tk1-5] (PROC GLIMMIX, SAS® 9.4; SAS Institute, Inc.). Tukey-Kramer adjustments controlled

205    Type I error inflation for multiple comparisons. The α-level for significance was $p$=0.05. We used robust

206    linear regression as implemented by the 'fitlm' function in MATLAB to assess links between neural and

207    behavioral measures.

**3.   RESULTS**

**3.1 Behavioral identification (%, RTs)**

210    Behavioral identification functions are shown across the different noise SNRs in Figure 3A.

211    Listeners' identification was more categorical (i.e., dichotomous) for clear speech and became more

212    continuous with poorer SNR. Analysis of the slopes ($\beta_1$) confirmed a main effect of SNR [$F_{2,28}$ =35.25,

213    $p$<0.0001] (Fig. 3B). Tukey-Kramer contrasts revealed psychometric slopes were unaltered for 0 dB SNR

214    relative to clear speech ($p$=0.33). However, -5 dB SNR noise weakened categorization, flattening the

215    psychometric function (-5dB vs. 0 dB, $p$<0.0001). These findings indicate the strength of categorical

216    representations is resistant to acoustic interference. That is, even when signal and noise compete at

217    equivalent levels, categorical processing persists. CP is weakened only for severely degraded speech (i.e.,

218    negative SNRs) where the noise exceeds the target signal.

219    Noise-related changes in the psychometric function could be related to uncertainty in category

220    distributions (prior probabilities) (Gifford et al., 2014) or lapses of attention due to task difficulty rather

221    than a weakening of speech categories, *per se* (Bidelman et al., under review). To rule out this latter

222    possibility, we used Bayesian inference (psignifit toolbox; Schütt et al., 2016) to estimate individual lapse

223    (λ) and guess (γ) rates from participants' identification data. Lapse rate (λ) was computed as the

7

224     difference between the upper asymptote of the psychometric function and 100%, reflecting the probability

225     of an "incorrect" response at infinitely high stimulus levels (i.e., responding "u" for Tk5; see Fig. 3A).

226     Guess rate ($\gamma$) was defined as the difference between the lower asymptote and 0. For an ideal observer

227     $\lambda$=0 and $\gamma$=0. We found neither lapse [$F_{2,28}$ =2.41, $p$=0.11] nor guess rate [$F_{2,28}$ =1.45, $p$=0.25] were

228     modulated by SNR. This helps confirm that while (severe) noise weakened CP for speech (Fig. 3B), those

229     effects were not driven by a lack of task vigilance or guessing, *per se* (Bidelman et al., under review;

230     Schütt et al., 2016).

231                                   [Insert Figure 3 near here]

232             The location of the perceptual boundary (Fig. 3C) varied marginally with SNR but the shift was

233     significant [$F_{2,28}$ =5.62, $p$=0.0089]. Relative to the clear condition, -5 dB SNR speech shifted the

234     perceptual boundary rightward ($p$=0.011). This indicates a small but measurable bias to report "u" (i.e.,

235     more frequent Tk1-2 responses) in the noisiest listening condition.

236             Behavioral RTs, reflecting the speed of categorization, are shown in Figure 3D. An ANOVA

237     revealed RTs were modulated by both SNR [$F_{2,200}$ =11.90, $p$<0.0001] and token [$F_{4,200}$ =5.36, $p$=0.0004].

238     RTs were similar when classifying clear and 0 dB SNR speech ($p$=1.0) but slowed in the -5 dB condition

239     ($p$<0.0001). Notably, *a priori* contrasts revealed this noise-related slowing in RTs was most prominent at

240     the phonetic endpoints of the continuum (Tk1-2 and Tk4-5); at the ambiguous (non-phonetic) Tk3, RTs

241     were identical across SNRs ($p$s > 0.69). This suggests that the observed RT effects in noise are probably

242     not due to a general slowing of decision speed (e.g., attentional lapses) across the board but rather, are

243     restricted to accessing categorical representations.

244             CP is also characterized by a slowing in RTs near the ambiguous midpoint of the continuum

245     (Bidelman et al., 2013; Bidelman & Walker, 2017; Bidelman et al., 2014b; Pisoni & Tash, 1974; Poeppel

246     et al., 2004; Reetzke et al., 2018). Planned contrasts revealed this characteristic slowing in RTs for the

247     clear [mean(Tk1,2,4,5) vs. Tk3; $p$=0.0003] and 0 dB SNR ($p$=0.0061) conditions. This categorical RT

248     pattern was not observed at -5 dB SNR ($p$=0.59). Collectively, our behavioral results suggest noise

249     weakened the strength of CP in both the quality and speed of categorical decisions but only when speech

250     was severely degraded. Perceptual access to categories was otherwise unaffected by low-level noise (i.e.,

251     $\geq$ 0 dB SNR).

252     **3.2 Electrophysiological data**

253             Grand average ERPs are shown across tokens and SNRs in Figure 4. ERPs showed the most

254     SNR- and token-related modulations starting around the P2 wave (~180 ms) that persisted for another 200

255     ms. Visual inspection of the data indicated these modulations were most prominent at centro-parietal

256    scalp locations. The enhanced positivity at these electrode sites following the auditory P2 might partly

257    reflect differences in P3b amplitude (Alain et al., 2001). To quantify these effects, we measured the mean

258    amplitudes in the 180-320 ms time window at the vertex channel (Cz) (Fig. 5). To assess the degree to

259    which ERPs showed categorical-level coding, we then pooled tokens Tk1 and Tk5 (those with clear

260    phonetic identities) and compared these responses to the ambiguous Tk3 at the midpoint of the continuum

261    (Bidelman, 2015; Bidelman & Walker, 2017). An ANOVA conducted on ERP amplitudes showed

262    responses were strongly modulated by SNR [$F_{2,70}$ =8.54, $p$=0.0005] and whether not the stimulus carried

263    a phonetic label [Tk1/5 vs. Tk3: $F_{1,70}$ =19.11, $p$ <0.0001] (Fig. 5B). Planned contrasts by SNR revealed

264    that neural activity differentiated phonetic vs. non-phonetic speech at clear ($p$=0.0170) and 0 dB

265    ($p$=0.0011) SNRs, but not at -5 dB ($p$=0.0915). Across SNRs, ERPs to phonetic tokens were largely

266    resilient to noise [Tk1/5; linear contrast of SNR: $t_{70}$ = -2.17, $p$=0.07)]. In contrast, responses declined for

267    non-phonetic speech sounds [Tk3; $t_{70}$ = -2.91, $p$=0.0098]. This indicates that that noise differentially

268    affected the representation of category prototypes relative to phonetically ambiguous speech. Pooled

269    across speech tokens, CLARA source analysis (Alain et al., 2017; Bidelman et al., 2018; Iordanov et al.,

270    2014) localized sensor-level activity to bilateral sources in anterior superior temporal gyrus (STG) (cf.

271    Bidelman & Lee, 2015). These neural findings parallel our behavioral results and suggest the categorical

272    (phonetic) representations of speech (near STG) are more resistant to noise than those that do not carry a

273    clear linguistic-phonetic identity.

274                                 [Insert Figure 4 near here]

275                                 [Insert Figure 5 near here]

276    **3.3 Brain-behavior relationships**

277         The effects of noise on categorical neural processing closely paralleled the perceptual data. Figure

278    6A shows the group mean performance on the behavioral identification task and group mean ERP

279    amplitudes (180-320 ms window) to the phonetic speech tokens (Tk1/5). For ease of comparison, both the

280    neural and behavioral measures were normalized for each participant (Alain et al., 2001), with 1.0

281    reflecting the largest displacement in ERP amplitude and psychometric slopes, respectively. The

282    remarkably similar pattern between brain and behavioral data imply that perceptual identification

283    performance is predicted by the underlying neural representations for speech, as reflected in the ERPs.

284    Indeed, correlational analyses revealed a strong association between behavioral responses and ERPs

285    elicited by the phonetic (Tk1/5) (Fig. 6B; $R^2$ = 0.15, $p$=0.0085) but not ambiguous (Tk3) tokens (Fig. 6C;

286    $R^2$ = 0.03, $p$=0.23). That is, more robust neural activity predicted steeper psychometric functions. These

287    findings suggest the neural processing of speech sounds carrying clear phonetic labels predicts more

288    dichotomous categorical decisions at the behavioral level; whereas neural responses to ambiguous (non-

289    categorical/non-phonetic) speech tokens do not predict perceptual categorization.

290                              [Insert Figure 6 near here]

291    **4. DISCUSSION**

292          By measuring neuroelectric activity during rapid classification of SIN, our results reveal three

293    main findings: (1) speech identification is robust to acoustic interference, degrading only at very severe

294    noise levels (i.e., negative SNRs); (2) the neural encoding of speech is enhanced for sounds carrying a

295    clear phonetic identity compared to phonetically ambiguous tokens; and (3) categorical neural

296    representations are more resistant to external noise than their non-categorical counterparts. Our findings

297    suggest the mere process of categorization—a fundamental operation to all perceptual systems (Goldstone

298    & Hendrickson, 2010)—aids figure-ground aspects of speech perception by fortifying abstract categories

299    from the acoustic signal and making the speech code more resistant to external noise interference.

300          Behaviorally, we found listeners' psychometric slopes were steeper when identifying clear

301    compared to noise-degraded speech; identification functions became shallower only at the severe

302    (negative) SNRs when noise levels exceeded that of speech. The resilience in perceptual identification

303    suggests the strength of categorical representations is largely resistant to signal interference.

304    Corroborating our modeling (Fig. 1), we found CP was affected only when the input signal was highly

305    impoverished. These data converge with previous studies (Bidelman et al., under review; Gifford et al.,

306    2014; Helie, 2017) suggesting category-level representations, which are by definition more abstract than

307    their acoustic-sensory counterparts, are largely impervious to surface degradations. Indeed, as

308    demonstrated recently in cochlear implant listeners, the sensory input can be highly impoverished, sparse

309    in spectrotemporal detail, and intrinsically noisy (i.e., delivered electrically to the cochlea) yet still offer

310    robust speech categorization (Han et al., 2016). Collectively, our data suggest that both the mere

311    construction of perceptual objects and the natural discrete binning process of CP help category members

312    to "pop out" amidst noise (e.g., Nothdurft, 1991; Perez-Gay et al., 2018) to maintain robust speech

313    perception in noisy environments.

314          Noise-related decrements in CP (Fig. 3A) could reflect a weakening of internalized categories

315    themselves (e.g., fuzzier match between signal and phonetic template) or alternatively, more general

316    effects due to task complexity (e.g., increased cognitive load or listening effort; reduced vigilance). We

317    can rule out the latter interpretation based on our RT data. The speed of listeners' perceptual judgments to

318    ambiguous speech tokens (Tk3) were nearly identical across conditions and invariant to noise (Fig. 3D).

319    In contrast, RT functions became more categorical ("inverted V" pattern) with increasing SNR due

320    entirely to changes in RTs for categorical members (continuum endpoints). These findings suggest that

10

321     categories represent local *enhancements* of processing within the normal acoustic space (e.g., Fig. 1)

322     which acts to sharpen categorical speech representations. That our data do not reflect gross changes in

323     task vigilance is further supported by two additional findings: (i) lapses in performance did not vary

324     across stimuli which suggests vigilance was maintained across conditions and (ii) ERPs predicted

325     behavioral CP only for speech sounds that carried clear phonetic categories (Fig. 6). Therefore, the effects

326     of noise on CP are most parsimoniously described as changes in the relative *sharpness* of the auditory

327     categorical boundary (Livingston et al., 1998). That is, under extreme noise, speech identification is

328     blurred, and the normal warping of the perceptual space is partially linearized, resulting in more

329     continuous speech identification.

330         On the basis of fMRI, Guenther et al. (2004) posited that the length of time auditory cortical cells

331     remain active after stimulus presentation might be shorter for category prototypes than for other sounds.

332     They further speculated "the brain may be reducing the processing time for category prototypes, rather

333     than reducing the number of cells representing the category prototypes (Guenther et al., 2004; p.55)."

334     Some caution is warranted when interpreting these results given the sluggishness of the fMRI BOLD

335     signal. Still, our data do not agree with Guenther et al. (2004)'s first assertion since ERPs showed larger

336     (enhanced) activations to categorical prototypes within 200 ms. However, our RT data do agree with their

337     second hypothesis. We found RTs were faster for prototypical speech (i.e., $RT_{Tk1/5} < RT_{Tk3}$) providing

338     confirmatory evidence that well-formed categories are processed more efficiently by the brain.

339         Our neuroimaging data revealed enhanced brain activity to phonetic (Tk1/5) relative to

340     perceptually ambiguous (Tk3) speech tokens. This finding indicates categorical-level processing occurs as

341     early as ~150-200 ms after sound arrives at the ear (Alho et al., 2016; Bidelman et al., 2013; Toscano et

342     al., 2018). Importantly, these results cannot be explained in terms of mere differences in exogenous

343     stimulus properties. On the contrary, endpoint tokens of our continuum were actually the most distinct in

344     terms of their acoustics. Yet, these prototypical stimuli elicited stronger neural activity than midpoint

345     tokens (i.e., Tk1/5 > Tk3), which was not attributable to trivial differences in signal SNR. These results

346     are broadly consistent with previous ERP studies (Altmann et al., 2014; Bidelman & Lee, 2015; Bidelman

347     et al., 2013; Bidelman et al., 2014a; Bidelman et al., 2014b; Dehaene-Lambertz, 1997; Phillips et al.,

348     2000), fMRI data (Binder et al., 2004; Kilian-Hütten et al., 2011), and near-field unit recordings (Bar-

349     Yosef & Nelken, 2007; Chang et al., 2010; Micheyl et al., 2005; Steinschneider et al., 2003), which

350     suggest auditory cortical responses code more than low-level acoustic features and reflect the early

351     formation of auditory-perceptual objects and abstract sound categories.

352         ERP effects related to CP (Fig. 5) were consistent with activity arising from the primary and

353     associative auditory cortices along the Sylvian fissure (Alain et al., 2017). The latency of these

11

354    modulations was comparable to our previous electrophysiological studies on CP (Bidelman & Alain,

355    2015b; Bidelman et al., 2013; Bidelman & Walker, 2017) and may reflect a modulation of the P2 wave.

356    P2 is associated with speech discrimination (Alain et al., 2010; Ben-David et al., 2011), sound object

357    identification (Leung et al., 2013; Ross et al., 2013), and the earliest formation of categorical speech

358    representations (Bidelman et al., 2013). Categorical neural enhancements (i.e., Tk1/5 > Tk3) also

359    persisted ~200 ms after P2, through what appeared to be a P3b-like deflection. Whether this wave reflects

360    a late modulation of P2 or a true P3b response is unclear, the latter of which is typically evoked in

361    oddball-type paradigms.

362        A similar "post-P2" wave (180-320 ms) has been observed during speech categorization tasks

363    (Bidelman & Alain, 2015b; Bidelman et al., 2013), which varied with perceptual (rather) than acoustic

364    classification. This response could represent integration or reconciliation of the input with a phonetic

365    memory template (Bidelman & Alain, 2015b) and/or attentional reorienting during stimulus evaluation

366    (Knight et al., 1989). Similar responses in this time window have also been reported during concurrent

367    sound segregation tasks requiring active perceptual judgments of the number and quality of auditory

368    objects (Alain et al., 2001; Alain et al., 2017; Bidelman & Alain, 2015a). The response might thus reflect

369    controlled processes covering a widely distributed neural network including medial temporal lobe and

370    superior temporal association cortices near parietal lobe (Alain et al., 2001; Dykstra et al., 2016). The

371    posterior scalp distribution of this late deflection is consistent with this interpretation (Fig. 4). Paralleling

372    the dynamics in our neural recordings, studies have shown that perceptual awareness of target signals

373    embedded in noise produces early focal responses between 100-200 ms circumscribed to auditory cortex

374    and posterolateral superior temporal gyrus that is followed by a broad, P3b-like response (starting ~300

375    ms) associated with perceived targets (Dykstra et al., 2016). It has been suggested this later response, like

376    the one observed here, is necessary to perceive target SIN or under the demands of higher perceptual load

377    (Dykstra et al., 2016; Gutschalk & Dykstra, 2014; Lavie et al.).

378        What might be the mechanism for categorical neural enhancements (i.e., $ERP_{Tk1/5} > ERP_{Tk3}$) and

379    their high flexibility in noise? In their experiments on categorical learning, Livingston et al. (1998)

380    suggested that when "category-relevant dimensions are not as distinctive, that is, when the boundary is

381    particularly 'noisy,' a mechanism for enhancing separation may be more readily engaged" (p. 742).

382    Phoneme category selectivity is observed early (<150 ms) (Alho et al., 2016; Bidelman et al., 2013;

383    Chang et al., 2010), particularly in left inferior frontal gyrus (pars opercularis) (Alho et al., 2016), but

384    only under active task engagement (Alho et al., 2016; Bidelman & Walker, 2017). While some nascent

385    form of categorical-like processing may occur pre-attentively (Bizley & Cohen, 2013; Chang et al., 2010;

386    Joanisse et al., 2007; Krishnan et al., 2009), it is clear that attention enhances the brain's ability to form

387    categories (Alho et al., 2016; Bidelman et al., 2013; Bidelman & Walker, 2017; Recanzone et al., 1993).

388    In animal models, perceptual learning leads to an increase in the size of cortical representation and
389    sharpening or tuning of auditory neurons for actively attended (but not passively trained) stimuli
390    (Recanzone et al., 1993). We recently demonstrated visual cues from a talker's face help sharpen sound
391    categories to provide more robust speech identification in noisy environments (Bidelman et al., under
392    review). While multisensory integration is one mechanism that can hone internalized speech
393    representations to facilitate CP, our data here suggest that goal-directed attention is another.

394        The neural basis of CP likely depends on a strong audition-sensory memory interface (Bizley &
395    Cohen, 2013; Chevillet et al., 2013; DeWitt & Rauschecker, 2012; Jiang et al., 2018) rather than cognitive
396    faculties, *per se* (attentional switching and IQ; Kong & Edwards, 2016). Moreover, the degree to which
397    listeners show categorical vs. gradient perception might reflect the strength of phonological processing,
398    which could have ramifications for understanding certain clinical disorders that impair sound-to-meaning
399    mapping (e.g., dyslexia; Calcus et al., 2016; Joanisse et al., 2000; Werker & Tees, 1987). CP deficits
400    might be more prominent in noise (Calcus et al., 2016). Thus, while relations between CP and language-
401    based learning disorders remains equivocal (Hakvoort et al., 2016; Noordenbos & Serniclaes, 2015), we
402    speculate that assessing speech categorization under the taxing demands of noise might offer a more
403    sensitive marker of impairment (e.g., Calcus et al., 2016).

404        More broadly, the noise-related effects observed here may account for other observations in the
405    CP literature. For example, cross-language comparisons between native and non-native speakers' CP
406    demonstrate language-dependent enhancements in native listeners in the form of steeper behavioral
407    identification functions (Bidelman & Lee, 2015; Iverson et al., 2003b; Xu et al., 2006) and more
408    dichotomous (categorical) neural responses to native speech sounds (Bidelman & Lee, 2015; Zhang et al.,
409    2011). Shallower categorical boundaries for non-native speakers can be parsimoniously described as
410    changes in *intrinsic* noise, which mirror the effects of *extrinsic* noise in the current study. While the noise
411    sources differ (exogenous vs. endogenous), both linearize the psychometric function and render speech
412    identification more continuous. Similarly, the introduction of visual cues of a talker's face can enhance
413    speech categorization (Bidelman et al., under review; Massaro & Cohen, 1983). Such effects have been
414    described as a reduction in decision ambiguity due to the mutual reinforcement of speech categories
415    provided by concurrent phoneme-viseme information (Bidelman et al., under review). Future studies are
416    needed to directly compare the impact of intrinsic vs. extrinsic noise on categorical speech processing.
417    Still, the present study provides a linking hypothesis to test whether deficits (Calcus et al., 2016; Joanisse
418    et al., 2000; Werker & Tees, 1987), experience-dependent plasticity (Bidelman & Lee, 2015; Xu et al.,
419    2006), and effects of extrinsic acoustics on CP (present study) can be described via common
420    physiological mechanisms.

13

421    **Conflicts of interest:** The authors declare no competing financial interests.

422

**Figure Legends**

**Figure 1: Theoretical framework for noise-related influences on categorical speech representations.**
**(A)** The neural representation of speech is modeled as a multidimensional feature space where populations of auditory cortical neurons code different dimensions (DIM) of the input. DIMS here are arbitrary but could reflect any behaviorally relevant feature of speech (e.g., F0, duration, etc.) Both 3D and 2D representations are depicted here for two stimulus classes. Categorical coding (modeled as a Gaussian mixture) is reflected by an increase in local firing rate for perceptually similar stimuli ("A" and "B"). **(B)** Noise blurs physical acoustic details yet spares categories as evidenced by the resilience of the peaks in neural space. Neural noise was modeled by changing the variance of additive Gaussian white noise.

**Figure 2: Acoustic spectrograms of the speech continuum as a function of SNR.** Vowel first formant frequency was parameterized over five equal steps (430 to 730 Hz, ►), resulting in a perceptual phonetic continuum from /u/ to /a/. Token durations were 100 ms. Speech stimuli were presented at 75 dB SPL with noise added parametrically to vary SNR.

**Figure 3: Behavioral speech categorization is robust to noise interference.** (**A**) Perceptual psychometric functions for clear and degraded speech identification. Curves show an abrupt shift in perception when classifying speech indicative of discrete perception (i.e., CP). (**B**) Slopes and (**C**) locations of the perceptual boundary show speech categorization is robust even down to 0 dB SNR; the strength of CP diminishes only for highly impoverished signals as indicated by the shallow slope and slight rightward bias of the curves (i.e., more frequently responding "u") in the -5 dB SNR condition. (**D**) Speech classification speeds (RTs) show a categorical pattern for clear and 0 dB SNR speech; participants are slower at labeling ambiguous tokens (midpoint) relative to those with a clear phonetic label (endpoints) (Bidelman & Walker, 2017; Pisoni & Tash, 1974). A categorical RT effect is not observed for highly degraded speech (-5 dB SNR). errorbars = ± s.e.m.

**Figure 4: Speech-ERPs as a function of speech token and noise (SNR).** Representative electrodes at central (Cz), temporal (T7/8) and parietal (Pz) scalp sites. Stimulus and noise-related modulations are most prominent between P2 and P3b (180-320 ms). (**A**) Phonetic speech tokens (Tk1, Tk5) elicit stronger ERPs than (**B**) ambiguous sounds without a clear category (Tk3). Noise weakens and prolongs the neural encoding of speech.

**Figure 5: Categorical neural organization limits the degradative effects of noise on cortical speech processing. (A)** Scalp auditory ERP waveforms (Cz electrode). (*shaded regions*) Stronger responses are

15

458    observed for phonetic exemplar vs. ambiguous speech tokens [i.e., mean(Tk1, Tk5) > Tk3] but this effect

459    varies with SNR. Brain volumes show Cortical Low resolution electromagnetic tomography Analysis

460    Recursively Applied (CLARA; BESA® v7) (Iordanov et al., 2014) functional maps projected onto the

461    BESA template brain (Richards et al., 2016) [latencies: 240 ms (clear), 272 ms (0 dB), 300 ms (-5 dB)].

462    CLARA activations (pooled across all speech tokens) localize ERP activity to anterior superior temporal

463    cortex bilaterally. As in sensor (scalp) responses (cf. Fig. 4), source activations diminish with increasing

464    noise (i.e., lower SNRs). (**B**) Mean ERP amplitude (180-320 ms window) is modulated by SNR and

465    phonetic status. Categorical neural encoding (Tk1/5 > Tk3) is observed for all but the noisiest listening

466    condition. errorbars = ± s.e.m. *$p<0.05$; ** $p<0.01$.

467    **Figure 6: Brain-behavior associations in categorical speech perception. (A)** Amplitudes of the

468    auditory ERPs overlaid with behavioral data (psychometric slopes). Neural and behavioral measures are

469    normalized for each participant (Alain et al., 2001), with 1.0 reflecting the largest displacement in ERP

470    amplitude (180-320 ms) and psychometric slopes, respectively. (**B-C**) Correlations between behavioral

471    CP and neural responses to (**B**) phonetic (Tk1/5) and (**C**) non-phonetic speech tokens (Tk3). Data are

472    aggregated across SNRs (n=45 observations). Behavioral CP is predicted only by neural activity to

473    phonetic tokens; larger ERP amplitudes elicited by Tk1/5 speech are associated with steeper, more

474    dichotomous CP. Solid lines, significant regression; Dotted lines, non-significant regression. ** $p<0.01$.

475

**References**

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance,* 27(5), 1072-1089.

Alain, C., Arsenault, J. S., Garami, L., Bidelman, G. M., & Snyder, J. S. (2017). Neural correlates of speech segregation based on formant frequencies of adjacent vowels. *Scientific Reports,* 7(40790), 1-11.

Alain, C., Campeanu, S., & Tremblay, K. (2010). Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *J Cogn Neurosci,* 22(2), 392-403.

Alain, C., McDonald, K., & Van Roon, P. (2012). Effects of age and background noise on processing a mistuned harmonic in an otherwise periodic complex sound. *Hearing Research,* 283, 126-135.

Alho, J., Green, B. M., May, P. J. C., Sams, M., Tiitinen, H., Rauschecker, J. P., et al. (2016). Early-latency categorical speech sound representations in the left inferior frontal gyrus. *NeuroImage,* 129, 214-223.

Altmann, C. F., Uesaki, M., Ono, K., Matsuhashi, M., Mima, T., & Fukuyama, H. (2014). Categorical speech perception during active discrimination of consonants and vowels. *Neuropsychologia,* 64C, 13-23.

Bar-Yosef, O., & Nelken, I. (2007). The effects of background noise on the neural responses to natural sounds in cat primary auditory cortex. *Frontiers in Computational Neuroscience,* 1(3), 1-14.

Ben-David, B. M., Campeanu, S., Tremblay, K. L., & Alain, C. (2011). Auditory evoked potentials dissociate rapid perceptual learning from task repetition without learning. *Psychophysiology,* 48(6), 797-807.

Bidelman, G. M. (2015). Induced neural beta oscillations predict categorical speech perception abilities. *Brain and Language,* 141, 62-69.

Bidelman, G. M. (2017). Communicating in challenging environments: Noise and reverberation. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay & A. N. Popper (Eds.), *Springer Handbook of Auditory Research: The frequency-following response: A window into human communication*). New York, N.Y.: Springer Nature.

Bidelman, G. M. (2018). Sonification of scalp-recorded frequency-following responses (FFRs) offers improved response detection over conventional statistical metrics. *Journal of Neuroscience Methods,* 293, 59-66.

Bidelman, G. M., & Alain, C. (2015a). Hierarchical neurocomputations underlying concurrent sound segregation: Connecting periphery to percept. *Neuropsychologia,* 68, 38-50.

Bidelman, G. M., & Alain, C. (2015b). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *Journal of Neuroscience,* 35(2), 1240 –1249.

Bidelman, G. M., Davis, M. K., & Pridgen, M. H. (2018). Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hearing Research,* 367, 149-160.

17

512 Bidelman, G. M., & Howell, M. (2016). Functional changes in inter- and intra-hemispheric auditory
513 cortical processing underlying degraded speech perception. *Neuroimage,* 124, 581-590.

514 Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech
515 in musicians and non-musicians. *Brain Research,* 1355, 112-125.

516 Bidelman, G. M., & Lee, C.-C. (2015). Effects of language experience and stimulus context on the neural
517 organization and categorical perception of speech. *Neuroimage,* 120, 191-200.

518 Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech
519 perception in the human auditory system. *NeuroImage,* 79(1), 201-212.

520 Bidelman, G. M., Sigley, L., & Lewis, G. (under review). Acoustic noise and vision differentially warp
521 speech categorization.

522 Bidelman, G. M., Villafuerte, J. W., Moreno, S., & Alain, C. (2014a). Age-related changes in the
523 subcortical-cortical encoding and categorical perception of speech. *Neurobiology of Aging,* 35(11), 2526-
524 2540.

525 Bidelman, G. M., & Walker, B. (2017). Attentional modulation and domain specificity underlying the
526 neural organization of auditory categorical perception. *European Journal of Neuroscience,* 45, 690-699.

527 Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014b). Coordinated plasticity in brainstem
528 and auditory cortex contributes to enhanced categorical speech perception in musicians. *European*
529 *Journal of Neuroscience,* 40, 2662-2673.

530 Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of
531 sensory and decision processes in auditory object identification. *Nature Neuroscience,* 7(3), 295-301.

532 Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature*
533 *Reviews Neuroscience,* 14(10), 693-707.

534 Calcus, A., Lorenzi, C., Collet, G., Colin, C., & Kolinsky, R. (2016). Is there a relationship between
535 speech identification in noise and categorical perception in children with dyslexia? *Journal of Speech,*
536 *Language, and Hearing Research,* 59(4), 835-852.

537 Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010).
538 Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience,* 13(11), 1428-
539 1432.

540 Chevillet, M. A., Jiang, X., Rauschecker, J. P., & Riesenhuber, M. (2013). Automatic phoneme category
541 selectivity in the dorsal auditory stream. *Journal of Neuroscience,* 33(13), 5208 –5215.

542 Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in
543 adults. *Neuroreport,* 8(4), 919-924.

544 DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream.
545 *Proceedings of the National Academy of Sciences of the United States of America,* 109(8), E505-514.

546 Dykstra, A. R., Halgren, E., Gutschalk, A., Eskandar, E. N., & Cash, S. S. (2016). Neural correlates of
547 auditory perceptual awareness and release from informational masking recorded directly from human
548 cortex: A case study. *Frontiers in Neuroscience,* 10, 472.

549    Gifford, A. M., Cohen, Y. E., & Stocker, A. A. (2014). Characterizing the impact of category uncertainty
550    on human auditory categorization behavior. *PLoS Computational Biology,* 10(7), e1003715.

551    Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews:*
552    *Cognitive Science,* 1(1), 69-78.

553    Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural
554    map formation. *Journal of the Acoustical Society of America,* 100(2 Pt 1), 1111-1121.

555    Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., & Tourville, J. A. (2004). Representation of sound
556    categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research,* 47(1), 46-57.

557    Gutschalk, A., & Dykstra, A. R. (2014). Functional imaging of auditory scene analysis. *Hearing*
558    *Research,* 307, 98-110.

559    Hakvoort, B., de Bree, E., van der Leij, A., Maassen, B., van Setten, E., Maurits, N., et al. (2016). The
560    role of categorical speech perception and phonological processing in familial risk children with and
561    without dyslexia. *Journal of Speech, Language, and Hearing Research,* 59(6), 1448-1460.

562    Han, J.-H., Zhang, F., Kadis, D. S., Houston, L. M., Samy, R. N., Smith, M. L., et al. (2016). Auditory
563    cortical activity to different voice onset times in cochlear implant users. *Clinical Neurophysiology,*
564    127(2), 1603-1617.

565    Harnad, S. R. (1987). *Categorical perception: The Groundwork of Cognition*. New York: Cambridge
566    University Press.

567    Helie, S. (2017). The effect of integration masking on visual processing in perceptual categorization.
568    *Brain and Cognition,* 116, 63-70.

569    Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and
570    second language acquisition. *Journal of the Acoustical Society of America,* 119(5), 3059–3071.

571    Hu, L., Mouraux, A., Hu, Y., & Iannetti, G. D. (2010). A novel approach for enhancing the signal-to-
572    noise ratio and detecting automatically event-related potentials (ERPs) in single trials. *NeuroImage,*
573    50(1), 99-111.

574    Iordanov, T., Hoechstetter, K., Berg, P., Paul-Jordanov, I., & Scherg, M. (2014). CLARA: classical
575    LORETA analysis recursively applied, *OHBM 2014*).

576    Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003a). A
577    perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition,* 87(1),
578    B47-57.

579    Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003b). A
580    perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition,* 87(1), 47-
581    57.

582    Jiang, X., Chevillet, M. A., Rauschecker, J. P., & Riesenhuber, M. (2018). Training humans to categorize
583    monkey calls: Auditory feature- and category-selective neural tuning changes. *Neuron,* 98(2), 405-
584    416.e404.

585    Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic
586    children: speech perception, phonology, and morphology. *Journal of Experimental Child Psychology,*
587    *77*(1), 30-60.

588    Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the
589    preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial
590    paradigm. *Cerebral Cortex, 17*(9), 2084-2093.

591    Kilian-Hütten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory cortex encodes the
592    perceptual interpretation of ambiguous sound. *Journal of Neuroscience, 31*(5), 1715-1720.

593    Knight, R. T., Scabini, D., Woods, D. L., & Clayworth, C. C. (1989). Contributions of temporal-parietal
594    junction to the human auditory P3. *Brain Research, 502*(1), 109-116.

595    Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue
596    weighting and executive function. *Journal of Phonetics, 59*, 40-57.

597    Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009). Experience-dependent neural
598    representation of dynamic pitch in the brainstem. *Neuroreport, 20*(4), 408-413.

599    Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes
600    of speech categories, monkeys do not. *Perception and Psychophysics, 50*(2), 93-107.

601    Lavie, N., Beck, D. M., & Konstantinou, N. Blinded by the load: attention, awareness and the role of
602    perceptual load. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1641),
603    20130205-20130205.

604    Leung, A. W., He, Y., Grady, C. L., & Alain, C. (2013). Age differences in the neuroelectric adaptation to
605    meaningful sounds. *PLoS One, 8*(7), e68892.

606    Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the
607    speech code. *Psychological Review, 74*(6), 431-461.

608    Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010).
609    Specialization along the left superior temporal sulcus for auditory categorization. *Cerebral Cortex,*
610    *20*(12), 2958-2970.

611    Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by
612    category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(3), 732-
613    753.

614    Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in
615    speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9*(5), 753-
616    771.

617    Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005). Perceptual organization of tone
618    sequences in the auditory cortex of awake macaques. *Neuron, 48*(1), 139 –148.

619    Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers:
620    Auditory processing or phonological coding? *Journal of Experimental Child Psychology, 64*(2), 199-231.

621    Myers, E. B., & Swan, K. (2012). Effects of category learning on neural sensitivity to non-native phonetic
622    categories. *Journal of Cognitive Neuroscience,* 24(8), 1695-1708.

623    Noordenbos, M. W., & Serniclaes, W. (2015). The categorical perception deficit in dyslexia: A meta-
624    analysis. *Scientific Studies of Reading,* 19(5), 340-359.

625    Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research,*
626    31(6), 1073-1078.

627    Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory.
628    *Neuropsychologia,* 9, 97-113.

629    Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and
630    ERP measurements. *Clinical Neurophysiology,* 112, 713-719.

631    Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise.
632    *Ear and Hearing,* 30(6), 653-661.

633    Perez-Gay, F., Sicotte, T., Theriault, C., & Harnad, S. (2018). Category learning can alter perception and
634    its neural correlate. *arXiv:1805.04619 [Preprint]*(Available from: https://arxiv.org/abs/1805.04619).

635    Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and
636    current density mapping. *Electroencephalogr Clin Neurophysiol,* 72(2), 184-187.

637    Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex
638    accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience,* 12(6),
639    1038-1055.

640    Picton, T. W., van Roon, P., Armilio, M. L., Berg, P., Ille, N., & Scherg, M. (2000). The correction of
641    ocular artifacts: A topographic perspective. *Clinical Neurophysiology,* 111(1), 53-65.

642    Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and
643    vowels. *Perception and Psychophysics,* 13(2), 253-260.

644    Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition,*
645    25(1-2), 21-52.

646    Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories.
647    *Perception and Psychophysics,* 15(2), 285-290.

648    Poeppel, D., Guillemin, A., Thompson, J., Fritz, J., Bavelier, D., & Braun, A. R. (2004). Auditory lexical
649    decision, categorical perception, and FM direction discrimination differentially engage left and right
650    auditory cortex. *Neuropsychologia,* 42(2), 183-200.

651    Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency
652    representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal
653    of Neuroscience,* 13(1), 87-103.

654    Reetzke, R., Xie, Z., Llanos, F., & Chandrasekaran, B. (2018). Tracing the trajectory of sensory plasticity
655    across different stages of speech learning in adulthood. *Current Biology,* 28(9), 1419-1427.e1414.

656 Richards, J. E., Sanchez, C., Phillips-Meek, M., & Xie, W. (2016). A database of age-appropriate average
657 MRI templates. *Neuroimage,* 124(Pt B), 1254-1259.

658 Ross, B., Jamali, S., & Tremblay, K. L. (2013). Plasticity in neuromagnetic cortical responses suggests
659 enhanced auditory object representation. *BMC Neurosci,* 14, 151.

660 Scherg, M., Ille, N., Bornfleth, H., & Berg, P. (2002). Advanced tools for digital EEG review: virtual
661 source montages, whole-head mapping, correlation, and phase analysis. *Journal of Clinical*
662 *Neurophysiology,* 19(2), 91-112.

663 Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian
664 estimation of psychometric functions for (potentially) overdispersed data. *Vision Research,* 122, 105-123.

665 Steinschneider, M., Fishman, Y. I., & Arezzo, J. C. (2003). Representation of the voice onset time (VOT)
666 speech parameter in population responses within primary auditory cortex of the awake monkey. *Journal*
667 *of the Acoustical Society of America,* 114(1), 307-321.

668 Toscano, J. C., Anderson, N. D., Fabiani, M., Gratton, G., & Garnsey, S. M. (2018). The time-course of
669 cortical responses to speech revealed by fast optical imaging. *Brain and Language,* 184, 32-42.
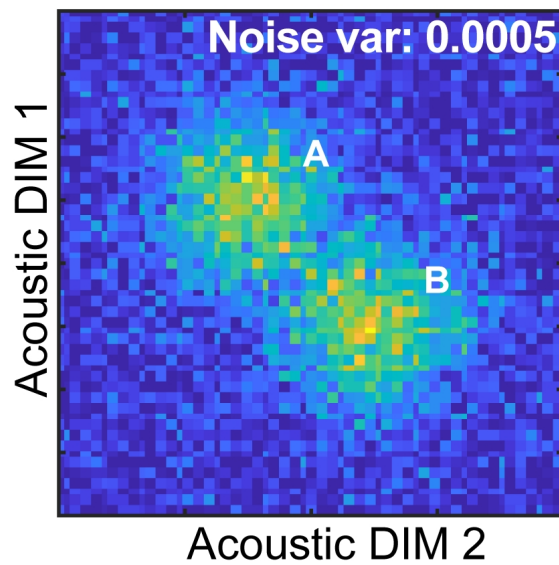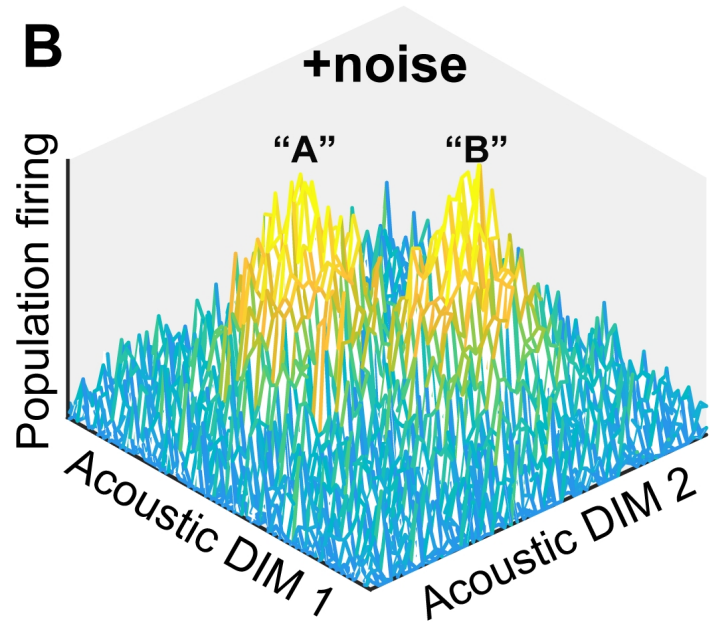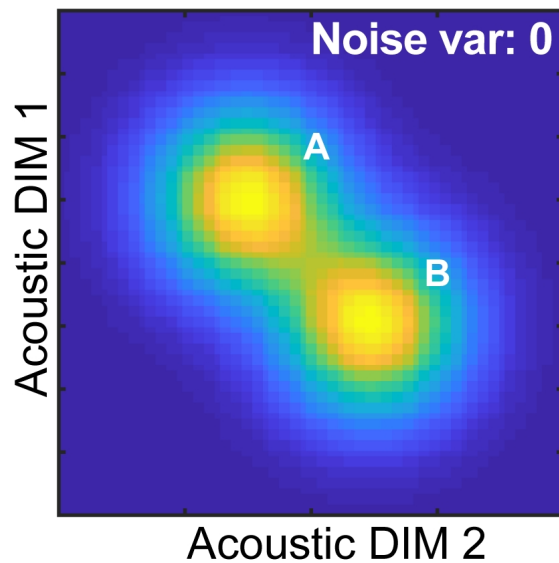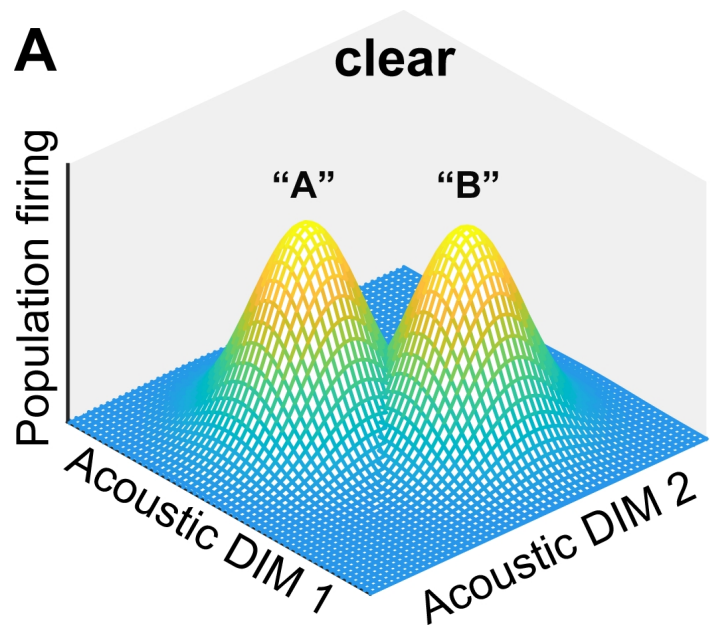
670 Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children.
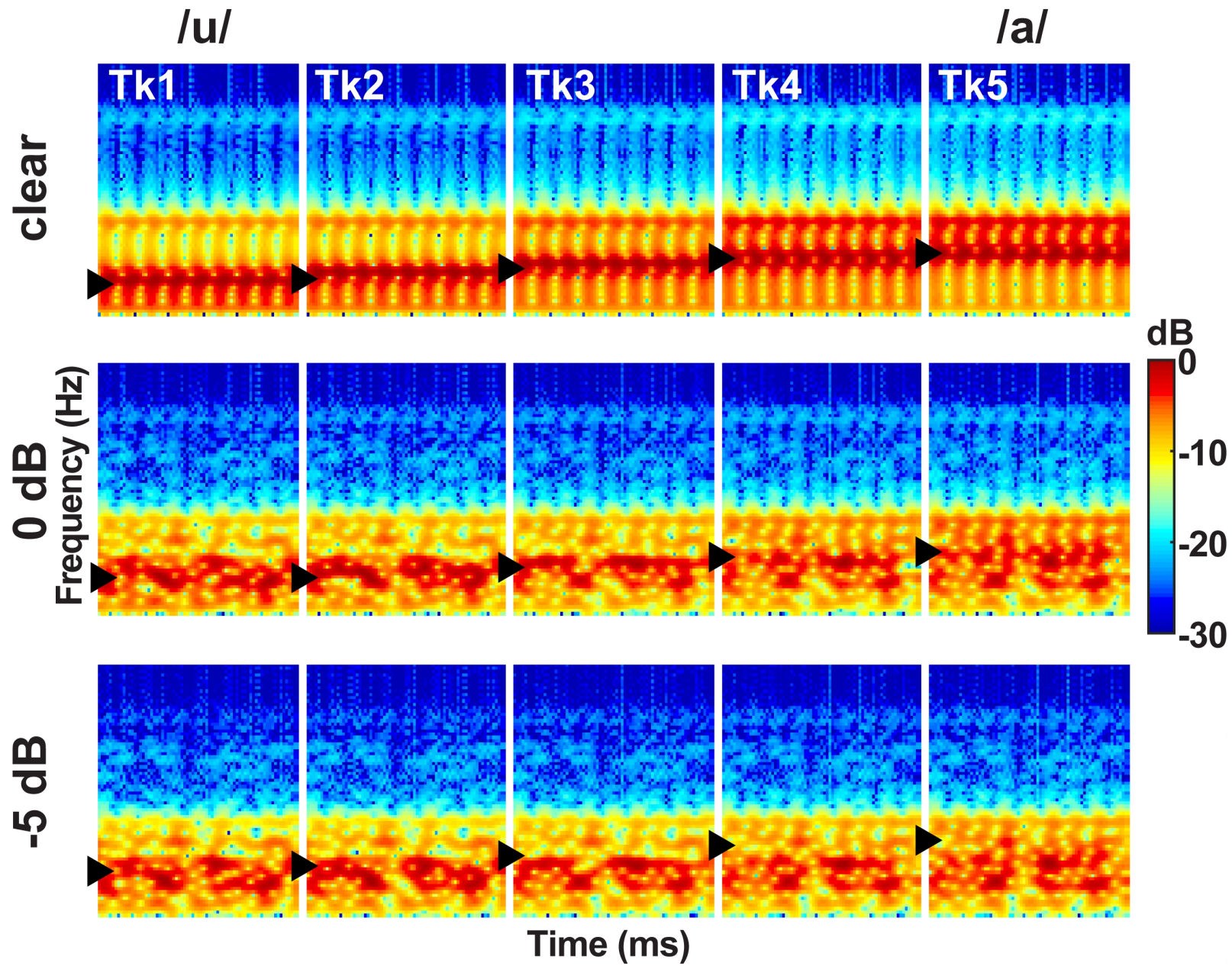671 *Canadian Journal of Psychology,* 41(1), 48-61.

672 Xu, Y., Gandour, J. T., & Francis, A. (2006). Effects of language experience and stimulus complexity on
673 the categorical perception of pitch direction. *Journal of the Acoustical Society of America,* 120(2), 1063-
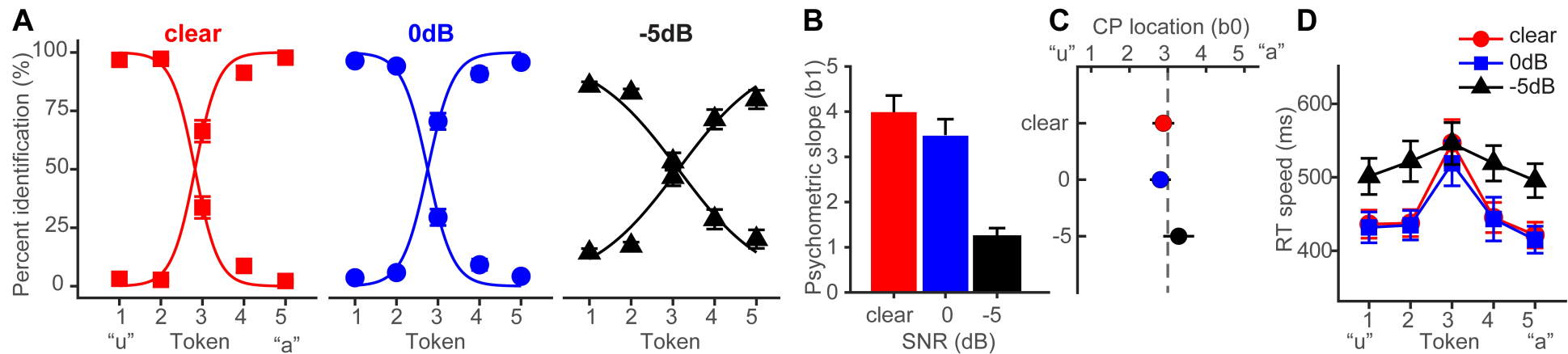674 1074.

675 Yoo, J., & Bidelman, G. M. (under review). Linguistic, perceptual, and cognitive factors underlying the
676 musician benefit to noise-degraded speech perception. *Hearing Research*.

677 Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., & Li, P. (2011). Cortical dynamics of acoustic and
678 phonological processing in speech perception. *PloS One,* 6(6), e20963.
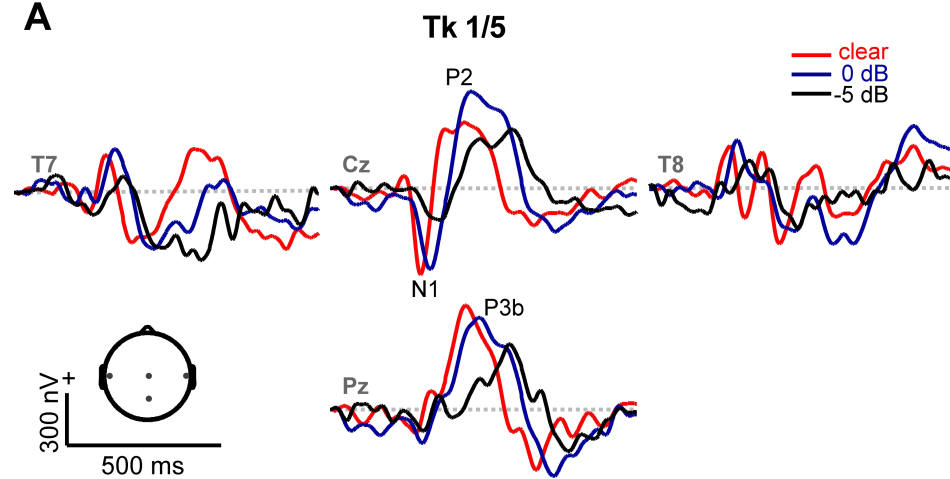
679

**A** clear
"A"  "B"
Population firing
Acoustic DIM 1
Acoustic DIM 2

Noise var: 0
Acoustic DIM 1
A
B
Acoustic DIM 2

**B** +noise
"A"  "B"
Population firing
Acoustic DIM 1
Acoustic DIM 2

Noise var: 0.0005
Acoustic DIM 1
A
B
Acoustic DIM 2

**A**

clear     0dB     -5dB

Percent identification (%)

100
75
50
25
0

1 "u"   2   3   4   5 "a"
Token

**B**

Psychometric slope (b1)

5
4
3
2
1
0

clear   0   -5
SNR (dB)

**C**

CP location (b0)

"u" 1   2   3   4   5 "a"

clear
0
-5

**D**

RT speed (ms)

600
500
400

1 "u"   2   3   4   5 "a"
Token

● clear
■ 0dB
▲ -5dB

**A**  Tk 1/5

**B**  Tk 3

T7   Cz   P2   T8

N1   P3b

Pz

clear
0 dB
-5 dB

300 nV
500 ms

**A**

clear

P1
P2
P3b?
N1
Cz

0 dB
Cz

-5 dB
Cz

0.25 μV +
200 ms

Tk1/5
Tk3

**B**

SNR
clear
0 dB
-5 dB

*

**

n.s.

ERP mean amplitude (μV) [180 -320 ms]

0.4

0.3

0.2

0.1

0

Tk1/5  Tk3          Tk1/5  Tk3          Tk1/5  Tk3

Speech token

**A**

**B** Phonetic tokens (Tk1/5)

**C** Non-phonetic tokens (Tk3)

R2=0.15**

R2=0.03, *n.s.*