



Lexical Influences on Categorical Speech Perception Are Driven by a Temporoparietal Circuit

Gavin M. Bidelman^{1,2}, Claire Pearson^{1*}, and Ashleigh Harrison^{1*}

Abstract

■ Categorical judgments of otherwise identical phonemes are biased toward hearing words (i.e., “Ganong effect”) suggesting lexical context influences perception of even basic speech primitives. Lexical biasing could manifest via late stage perceptual mechanisms related to decision or, alternatively, top-down linguistic inference that acts on early perceptual coding. Here, we exploited the temporal sensitivity of EEG to resolve the spatiotemporal dynamics of these context-related influences on speech categorization. Listeners rapidly classified sounds from a /g/-/k/ gradient presented in opposing word–nonword contexts (*GIFT–kift* vs. *giss–KISS*), designed to bias perception toward lexical items. Phonetic perception shifted toward the

direction of words, establishing a robust Ganong effect behaviorally. ERPs revealed a neural analog of lexical biasing emerging within ~200 msec. Source analyses uncovered a distributed neural network supporting the Ganong including middle temporal gyrus, inferior parietal lobe, and middle frontal cortex. Yet, among Ganong-sensitive regions, only left middle temporal gyrus and inferior parietal lobe predicted behavioral susceptibility to lexical influence. Our findings confirm lexical status rapidly constrains sublexical categorical representations for speech within several hundred milliseconds but likely does so outside the purview of canonical auditory-sensory brain areas. ■

INTRODUCTION

An important building block for language is the ability to transform sensory information into abstract linguistic representations (Goldstone & Hendrickson, 2010). Speech sounds vary continuously across time, environments, speaker identities, and stimulus contexts, and yet, listeners easily parse the speech stream into discrete phonemes (Lotto & Holt, 2016; Phillips, 2001; Pisoni & Luce, 1987). The categorical perception (CP) of speech maps infinitely variable acoustic signals into discrete phonetic–linguistic representations on which the speech-language system can operate (Pisoni & Luce, 1987; Pisoni, 1973; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). CP is indicated when gradually morphed speech sounds along a continuum are heard as belonging to one of a few discrete phonetic classes. Tokens labeled with different identities are said to cross the categorical boundary, a psychological border where listeners’ responses abruptly flips because of a perceptual warping of the stimulus space (i.e., compression of within-category sounds; Best & Goldstone, 2019; Goldstone, Steyvers, Spencer-Smith, & Kersten, 2000; Livingston, Andrews, & Harnad, 1998).

One nebulous issue in speech perception concerns whether higher-level activation of lexical representations directly affects sublexical components (e.g., phoneme categories). On one extreme is the rigid view that, once

established, internalized speech prototypes (i.e., equivalence classes or category members) are invariant to superficial stimulus manipulation or lexical context (Liberman, Harris, Hoffman, & Griffith, 1957). Under this model, categories are impervious to influences from surrounding information and sound elements that precede or follow an isolated stimulus cannot influence its categorization or location of the perceptual boundary. On the contrary, acoustic–phonetic categories—traditionally considered early or lower-level constructs of the speech signal—are in fact highly malleable to contextual variations (Holt & Lotto, 2010; Myers & Blumstein, 2008; Francis & Ciocca, 2003; Norris, McQueen, & Cutler, 2003; Elman & McClelland, 1988; Ganong, 1980; Pisoni, 1975). Moreover, the degree to which context influences the category identity of speech varies with language experience (Bidelman & Lee, 2015; Lively, Logan, & Pisoni, 1993; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Consequently, it is now well-established that phonetic categories are flexible and perception of even individual speech features depends critically on the surrounding signal (Repp & Liberman, 1987).

Context-dependent effects in CP are best illustrated by the so-called “Ganong effect” (Ganong, 1980). The Ganong phenomenon occurs when listeners’ perceived category boundary of a word–nonword continuum of phonemes shifts (is biased) toward the lexical item. When perceiving a “da-ta” continuum, for example, English-speaking listeners show a stark shift in their perceptual category boundary toward lexical items when one of the gradient’s endpoints contains a real word (e.g., “DASH-tash”;

¹University of Memphis, ²University of Tennessee Health Sciences Center, Memphis

*These authors contributed equally to this work.

Ganong, 1980; Ganong & Zatorre, 1980). Similar interpretive biasing can be induced via learning when listeners are exposed to new contexts that shape their perception of otherwise isolated sounds (Norris et al., 2003). Collectively, behavioral studies suggest that stimulus context expands the mental category for expected or behaviorally relevant stimuli (McMurray, Dennhardt, & Struck-Marcell, 2008).

One interpretation of lexical effects is that they reflect direct linguistic influence on perceptual processes. Alternatively, another school of thought argues lexical context effects are postperceptual and are therefore related to executive mechanisms (i.e., response selection, decision). Fox (1984) tested the interaction between lexical knowledge and phonetic categorization during speech perception using Ganong-like stimuli. Lexical status did not influence phonetic categorization at shorter response latencies or when participants were given a response deadline, suggesting lexical context influences later stimulus selection rather than perceptual encoding, *per se*. This notion is supported by results from Pitt and Samuel (1993), who found the strength of lexical influences on perception of ambiguous sound tokens depended on their position in a word; lexical effects were weaker when tokens occurred toward the beginning compared to the end of words. These data support “late stage” or “selection-based” models whereby the very formation of categories themselves only emerges at a late decision stage of the processing hierarchy (e.g., MERGE model; Norris, McQueen, & Cutler, 2000).

Rather than acting at late stages, lexical biasing could instead manifest via top-down (and perhaps bi-directional) modulations of early perceptual processing with the lexical interface. Indeed, growing evidence from neuroimaging studies (Noe & Fischer-Baum, 2020; Gow, Segawa, Ahlfors, & Lin, 2008; Myers & Blumstein, 2008; van Linden, Stekelenburg, Tuomainen, & Vroomen, 2007) reaffirms such interactive, connectionist views of categorization (e.g., TRACE; McClelland & Elman, 1986). Employing fMRI with a Ganong task, Myers and Blumstein (2008) found that the placement of the phonetic boundary modulated activity both in perceptual (e.g., superior temporal gyrus [STG]), inferior parietal lobe [IPL]) and frontal executive brain areas (inferior frontal gyrus, ACC), with greater activity for ambiguous items near the boundary. The mere involvement of the STG strongly suggests that lexical shifts are not solely due to executive decision processes but, at minimum, includes a perceptual component that either itself has direct access to lexical properties or is interactively reactivated to integrate phonetic and extraphonetic factors in placing the phonetic boundary (Noe & Fischer-Baum, 2020; Gow et al., 2008; Myers & Blumstein, 2008). Although fMRI offers excellent spatial characterization of potential lexical effects, it lacks the temporal precision necessary to resolve the underlying brain dynamics of category formation (Bidelman, Moreno, & Alain, 2013) and related lexical influences (Gow et al., 2008), both of which unfold within a few hundred milliseconds after speech onset (e.g., Mahmud, Yeasin, & Bidelman, 2020).

Extending prior neuroimaging work (Gow et al., 2008; Myers & Blumstein, 2008), the aim of this study was to characterize the spatiotemporal dynamics of context-dependent lexical influences on CP with the goal of establishing where and when speech categories are prone to Ganong-like biasing. We used EEG coupled with source reconstruction to assess the underlying neural bases of phoneme categorization and its lexical modulation. Our task included word-nonword (*GIFT-kift*) and nonword-to-word (*giss-KISS*) acoustic gradients of an otherwise identical /g₁-/k₁/ acoustic-phonetic continuum designed to bias listeners’ perception toward the lexical item and shift their perceptual category boundary (Myers & Blumstein, 2008; Ganong, 1980). Our findings confirm that lexical status rapidly (~200–300 msec) constrains sublexical category speech representations but further suggests this interactivity occurs outside canonical auditory-linguistic brain structures. Instead, among Ganong-sensitive brain regions, we find engagement of a temporoparietal circuit (i.e., inferior parietal, middle temporal gyrus [MTG]) is critical to describing listeners’ susceptibility to contextual biasing during category judgments.

METHODS

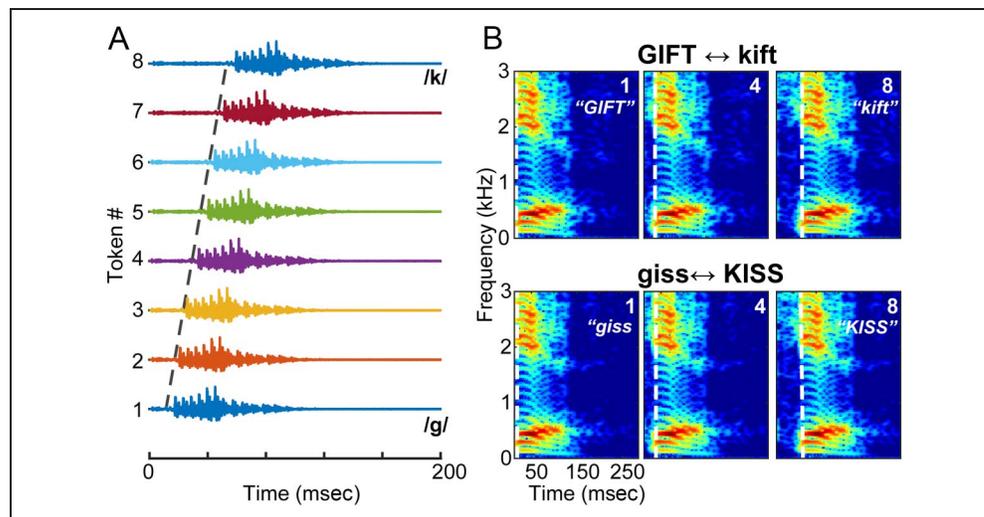
Participants

Sixteen young adults (3 men, 13 women; age: $M = 24.5$, $SD = 12.9$ years) were recruited from the University of Memphis student body.¹ Sample size was based on several previous neuroimaging studies on context effects in CP (e.g., Gow et al., 2008; Myers & Blumstein, 2008). All exhibited normal hearing sensitivity confirmed via audiometric screening (i.e., < 25 dB HL, octave frequencies 250–8000 Hz). Each participant was strongly right-handed ($74.8 \pm 27.0\%$ laterality index; Oldfield, 1971), had obtained a collegiate level of education (18.8 ± 2.7 years formal schooling), and was a native speaker of American English. Participants were considered nonmusicians (e.g., Mankel & Bidelman, 2018), having, on average, 3.25 ± 3.3 years of music training. All were paid for their time and gave informed consent in compliance with a protocol approved by the institutional review board at the University of Memphis.

Speech Stimulus Continua

Stimuli were adapted from Myers and Blumstein (2008). Speech tokens consisted of a /g₁/ to /k₁/ (i.e., “gih” to “kih”) stop-consonant continuum presented in two word/nonword contexts.² Each continuum was constructed using eight equally spaced VOTs incrementing from 18 msec (/g/ percept) to 70 msec (/k/ percept; Figure 1). This otherwise identical VOT continuum was used to create word-to-nonword (*GIFT-kift*) and nonword-to-word (*giss-KISS*) gradients designed to bias listeners’ phonemic perception toward the lexical item (Figure 1B). This was achieved by splicing the appropriate aspiration (i.e., “-ft” for

Figure 1. Speech stimuli used to probe the neural basis of lexical effects on categorical speech processing. (A) Acoustic waveforms of the continuum (zoomed to 200 msec). Stimuli varied continuously in equidistant VOT steps to yield a morphed gradient from /gi/ to /ki/. (B) Spectrograms. The /gi/ to /ki/ continuum was presented in one of two word–nonword contexts (*GIFT–kift* and *giss–KISS*) such that, at any point along the acoustic gradient, the same stop consonant could be perceived more as a word (or nonword) depending on lexical bias from the continuum’s endpoint. Dotted lines, onset of voicing demarcating VOT duration.



GIFT–kift; “-ss” for *giss–KISS*) to the end of the otherwise identical /gi-/ki/ sounds (for details, see Myers & Blumstein, 2008). All tokens were 500 msec in duration and root-mean-square amplitude normalized.

During EEG recording, listeners heard 120 trials of each individual token (per context) in which they labeled the sound with a binary response (“g” or “k”) as quickly and accurately as possible. Following, the ISI was jittered randomly between 800 and 1000 msec (20-msec steps, uniform distribution) to avoid rhythmic entrainment of the EEG and anticipating subsequent stimuli. Block order for the *GIFT–kift* versus *giss–KISS* continua were randomized within and between participants. The auditory stimuli were delivered binaurally at 79 dB SPL through shielded insert earphones (ER-2; Etymotic Research) controlled by a TDT RP2 signal processor (Tucker Davis Technologies).

EEG Recordings

EEGs were recorded from 64 sintered Ag/AgCl electrodes at standard 10–10 scalp locations (Oostenveld & Praamstra, 2001). Continuous data were digitized at 500 Hz (SynAmps RT amplifiers; Compumedics Neuroscan) using an online passband of DC–200 Hz. Electrodes placed on the outer canthi of the eyes and the superior and inferior orbit monitored ocular movements. Contact impedances were maintained < 10 kΩ. During acquisition, electrodes were referenced to an additional sensor placed ~ 1 cm posterior to Cz. Data were rereferenced off-line to the common average for analysis. Preprocessing was performed in BESA Research (v7.1; BESA, GmbH). Ocular artifacts (saccades and blinks) were corrected in the continuous EEG using PCA (Picton et al., 2000). Cleaned EEGs were then filtered (1–20 Hz), epoched (–200 to 800 msec), baselined to the prestimulus interval, and ensemble averaged resulting in 16 ERP waveforms per participant (8 tokens × 2 contexts).

Behavioral Data Analysis

Identification scores were fit with a sigmoid function $P = 1/[1 + e^{-\beta_1(x - \beta_0)}]$, where P is the proportion of trials identified as a given phoneme, x is the step number along the stimulus continuum, and β_0 and β_1 the location and slope of the logistic fit estimated using nonlinear least-squares regression. Comparing parameters between speech contexts revealed possible differences in the “steepness” (i.e., rate of change) and, more critically, the location of the categorical boundary as a function of speech context. A lexical bias (i.e., Ganong effect) is indicated when the location of the perceptual boundary (β_0) in phoneme identification shifts dependent on the anchoring speech context (Myers & Blumstein, 2008; Ganong, 1980). Behavioral labeling speeds (i.e., RTs) were computed as listeners’ median response latency across trials for a given condition. RTs outside 250–2500 msec were deemed outliers (e.g., fast guesses, lapses of attention) and were excluded from the analysis (Bidelman et al., 2013; Bidelman & Walker, 2017).

EEG Data Analysis

ERP Sensor Responses

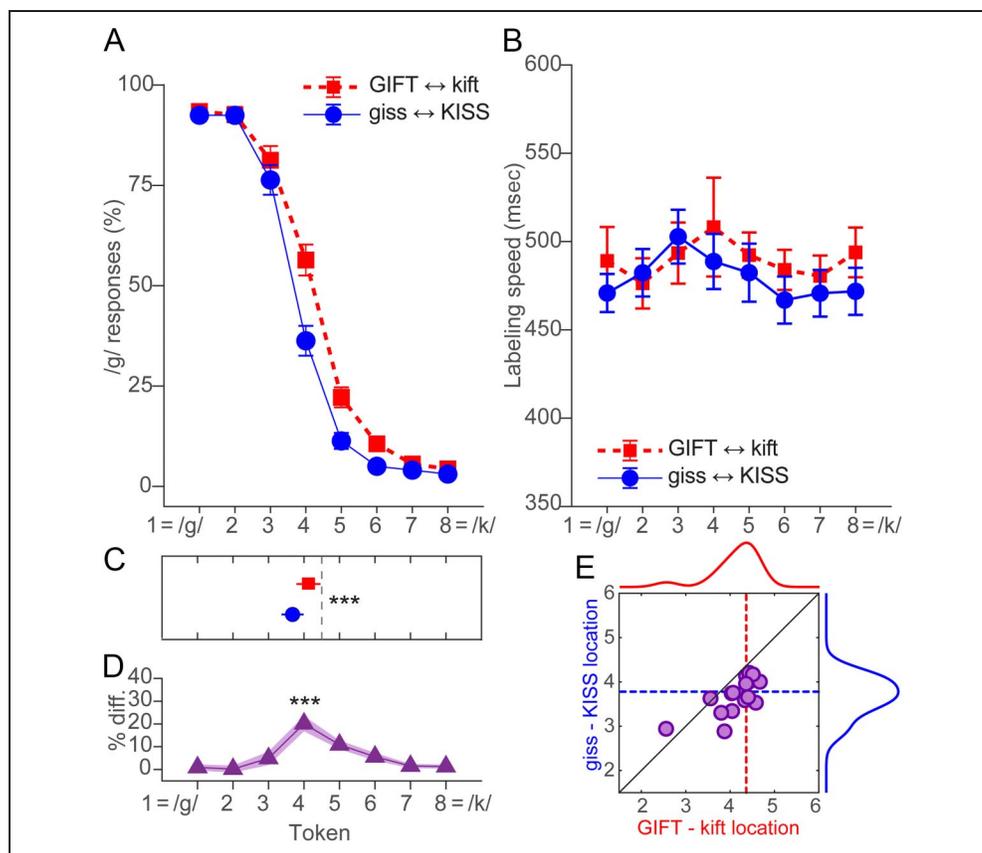
From channel-level waveforms, we measured lexical bias effects in the speech ERPs by comparing scalp topographies at the ambiguous midpoint token (Tk4) evoked in the two different speech contexts (i.e., *GIFT*4 vs. *KISS*4). This token step is where lexical bias effects were most prominent behaviorally (see Figure 2). Topographic t tests were conducted in EEGLAB (Delorme & Makeig, 2004).

Source Analysis

To estimate the underlying sources contributing to the lexical effect, we used Classical Low Resolution Electromagnetic

Figure 2. Lexical context biases the perceptual categorization of speech. (A) Psychometric identification functions show a shift in the perceptual boundary toward lexical items. Listeners more frequently reported /g/ responses in the *GIFT*–*kift* continuum and more /k/ responses for the *giss*–*KISS* context, confirming perception for otherwise identical stop consonants is biased toward hearing words. (B) RTs. Labeling speeds are faster for endpoint versus midpoint tokens of the continuum consistent with category ambiguity near the midpoint of the continuum (Pisoni & Tash, 1974). (C) Critically, the location of the perceptual boundary (i.e., β_0) shifts depending on the lexical context. (D) Identification performance differs maximally between contexts near the midpoint of the continua (i.e., Tk 4). (E) Comparison of boundary locations (β_0) for the *GIFT*–*kift* versus *giss*–*KISS* continua. The diagonal represents the case of an identical perceptual boundary

between contexts. Boundaries shift leftward for *giss*–*KISS* compared to *GIFT*–*kift*, reflecting a higher precedence of /k/ responses in that context (vice versa for the other context). Error bars = ± 1 SEM; *** $p < .0001$.



Tomography Analysis Recursively Applied (CLARA; BESA (v7); Iordanov, Hoechstetter, Berg, Paul-Jordanov, & Scherg, 2014) to estimate the neuronal current density underlying the scalp ERPs (e.g., Bidelman, 2018; Alain, Arsenault, Garami, Bidelman, & Snyder, 2017). CLARA models the inverse solution as a large collection of elementary dipoles distributed over nodes on a mesh of the cortical volume. The algorithm estimates the total variance of the scalp data and applies a smoothness constraint to ensure current changes minimally between adjacent brain regions (Michel et al., 2004; Picton et al., 1999). CLARA renders more focal source images by iteratively reducing the source space during repeated estimations. On each iteration ($\times 2$), a spatially smoothed LORETA solution (Pascual-Marqui, Esslen, Kochi, & Lehmann, 2002) was recomputed and voxels below a 1% max amplitude threshold were removed. This provided a spatial weighting term for each voxel on the subsequent step. Two iterations were used with a voxel size of 7 mm in Talairach space and regularization (parameter accounting for noise) set at 0.01% singular value decomposition. Source activations were visualized on BESA's adult brain template (Richards, Sanchez, Phillips-Meek, & Xie, 2016).

To quantify the time course of source activations, we seeded discrete dipoles within the activation centroids

identified in the CLARA volume images at a latency of 286 msec, where scalp data showed maximally lexical effects (see Figure 4A). CLARA localized activity to five major foci including MTG, inferior parietal lobe (IPL), and middle frontal gyrus (MFG) in left hemisphere, and precentral gyrus (PrCG) and insular cortex (IC) of right hemisphere (see Figure 4D). Dipole time courses represent the estimated current within each regional source. We then used this 5-dipole model to create a virtual source montage to transform each participant's scalp potentials (sensor-level recordings) into source space (Scherg, Berg, Nakasato, & Beniczky, 2019; Scherg, Ille, Bornfleth, & Berg, 2002). This digital remontaging applied a spatial filter to all electrodes (defined by the foci of our dipole configuration) to transform the electrode recordings to a reduced set of source signals reflecting the neuronal current (in units nAm) as seen within each anatomical ROI (Bidelman, 2018; Bidelman, Davis, & Pridgen, 2018). Critically, we fit individual dipole orientations to each participant's own data (anatomical locations remained fixed) to maximize the explained variance of the model at the individual subject level. The model provided a good fit to the grand averaged scalp data (goodness of fit, entire epoch window = 75%), confirming the ERPs could be described by a restricted number of sources.

Brain–Behavior Correspondence

From the source waveform time courses, we measured peak amplitudes within the 200- to 300-msec time window, where lexical effects were prominent in raw EEG data (see Figure 4A, B). We then regressed source amplitudes (for each ROI) with listeners' behavioral Ganong effect, computed as the magnitude of shift in their perceptual boundary between speech contexts (i.e., data in Figure 2C). This allowed us to assess the behavioral relevance of each brain ROI and how context-dependent changes in neural activity (i.e., “neural Ganong” effect) relate to lexical biases in CP measured behaviorally.

Statistics

We analyzed the data using mixed-model ANOVAs in R (R Core team, 2018; *lmer4* package) with fixed effects of token (eight levels) and speech context (two levels). Participants served as a random effect. Multiple comparisons were corrected using Tukey–Kramer adjustments. Brain–behavior relations were assessed using robust regression (bisquare weighting) performed using the *fitlm* function in MATLAB 2020a (The MathWorks, Inc.). Effect sizes are reported for omnibus ANOVAs using Cohen's *d* (Cohen, 1988), for paired *t* tests using the formula described in Dunlap, Cortina, Vaslow, and Burke (1996), and as Pearson's *r* for correlations.

RESULTS

Behavioral Data

Behavioral identification functions are shown for the two speech contexts in Figure 2A. Listeners more frequently reported /g/ responses in the *GIFT–kift* continuum and more /k/ responses for the *giss–KISS* context, confirming that perception for otherwise identical stop consonants is biased toward hearing words. The perceptual boundary location depended strongly on context, $t(15) = 4.82, p < .0001; d = 0.961$ (Figure 2C and 2E). Consistent with prior studies (Noe & Fischer-Baum, 2020; Myers & Blumstein, 2008; Ganong, 1980), context-dependent effects in CP were most evident near the ambiguous midpoint of the continuum (Tk 4), where listeners' identification abruptly shifted phoneme categories, $t(15) = 6.00, p < .0001; d = 2.19$ (Figure 2D). Ganong shifts also varied across individuals (e.g., Lam, Xie, Tessmer, & Chandrasekaran, 2017), with some listeners showing strong influence to lexical bias and others showing little to no changes in perception with speech context (Figure 3).

Speech labeling speeds were modulated by context, $F(1, 225) = 5.15, p = .024; d = 0.270$, and token, $F(7, 225) = 2.14, p = .0408; d = 0.370$, (Figure 2B). Identification was faster overall when categorizing tokens in the *giss–KISS* context ($p = .024$). The main effect of token was attributable to a slowing of RTs near the midpoints of each continua (i.e., mean[1278] vs. mean[45] contrast: $t(225) =$

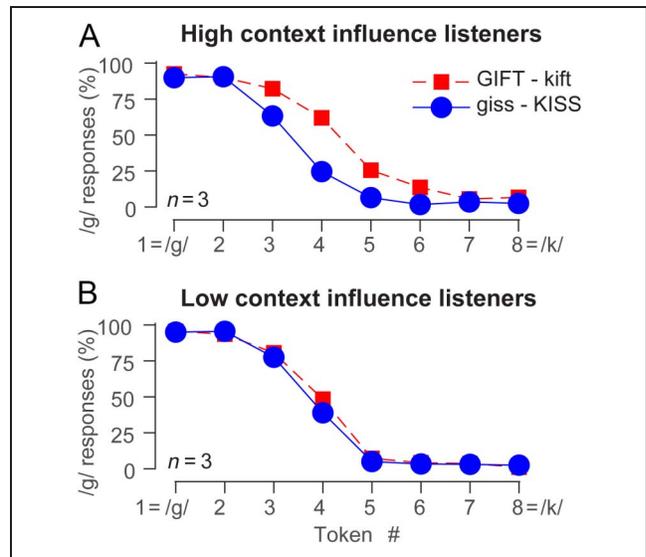


Figure 3. Lexical influences on CP are subject to individual differences. Identification functions for representative listeners ($n = 3$) who showed the strongest (A) and weakest (B) influence of lexical context on speech categorization. High influence listeners' perceptual boundary shifts dramatically with context, whereas low influence listeners show little change in perception with lexical context.

3.14, $p = .0019$). Such inverted V shape in labeling speeds, although not prominent in these data, have been attributed to more ambiguity in decision nearer the perceptual boundary (Bidelman & Walker, 2017; Pisoni & Tash, 1974). Collectively, these behavioral results suggest that lexical information (words) biases listeners' categorization of otherwise identical phonetic features; even basic phoneme perception is latticed by the surrounding lexical context of the speech signal.

EEG Data

Scalp ERPs are shown at electrode Cz in Figure 4. To quantify the “neural Ganong” effect, we contrasted ERPs to tokens at the perceptual boundary (i.e., Tk 4; e.g., Myers & Blumstein, 2008), where lexical bias was strongest behaviorally (see Figure 2). Difference waves computed between midpoint tokens evoked during *giss–KISS* versus *GIFT–kift* continua revealed context-dependent modulations in the time window between 200–300 msec, $t(14) = 3.03, p = .009; d = 1.15$ (Figure 4A and 4B).³ That is, despite identical acoustic information, phonemes were processed differentially depending on the word context they carried. The topography of the neural Ganong was broadly distributed over the scalp, spanning frontal, temporal, and parietal electrodes (Figure 4C).

ERP differences between ki(ss) and gi(ft) could be because of lexical biasing of the initial phoneme or the fact that boundary tokens carry different word endings. That is, for stimuli near the category boundary, one token is a real word whereas the other is equivocal in lexical status. To rule out this possibility, after Myers and Blumstein (2008), we

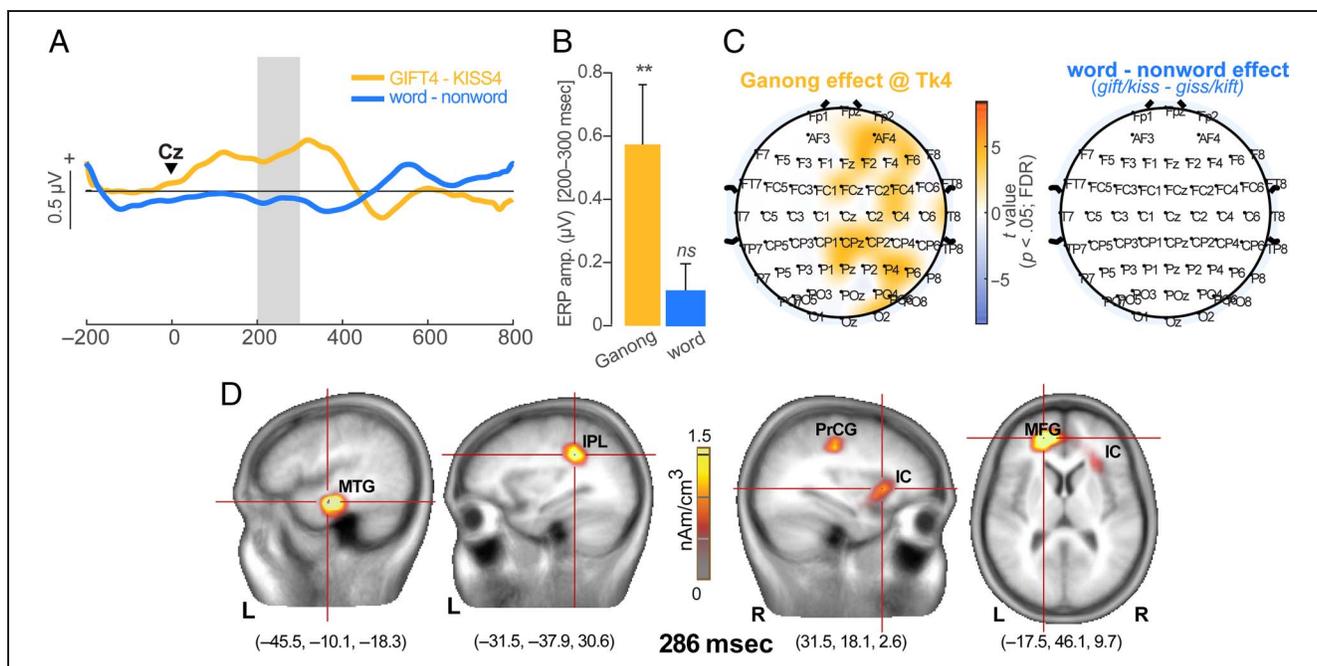


Figure 4. Neuroelectric brain activity reveals evidence of lexical biasing on speech categories. (A) ERP time course at the Cz electrode reflecting difference waves between Tk 4 responses when presented in *GIFT*–*kift* versus *giss*–*KISS* contexts (i.e., lexical effect contrast) and words (*gift*–*kiss*) versus nonwords (*giss*–*kift*; i.e., word contrast). \blacktriangledown = speech stimulus onset. A running *t* test (Guthrie & Buchwald, 1991) reveals lexical biasing between 200 and 300 msec ($p < .05$; shaded segment). (B) Mean ERP amplitude at Cz (200–300 msec) differs from 0 for the Ganong (but not word control) contrast indicating differentiation of identical speech tokens dependent on lexical context. (C) Topographic distribution of the Ganong effect across the scalp. Statistical maps (paired *t* test, $p < .01$, FDR-corrected; Benjamini & Hochberg, 1995) contrasting Tk4 responses in the two lexical contexts. No clusters emerged in the word–nonword contrast suggest that Ganong biasing is not because of the word status of stimuli, per se. (D) Brain volumes show CLARA (Iordanov et al., 2014) distributed source activation maps underlying lexical bias during speech categorization. Maps were rendered at latency of 286 msec, where the effect was most prominent at the scalp (e.g., Figure 4A). Functional data are overlaid on an adult template brain (Richards et al., 2016). Error bars = ± 1 SEM; ** $p < .01$.

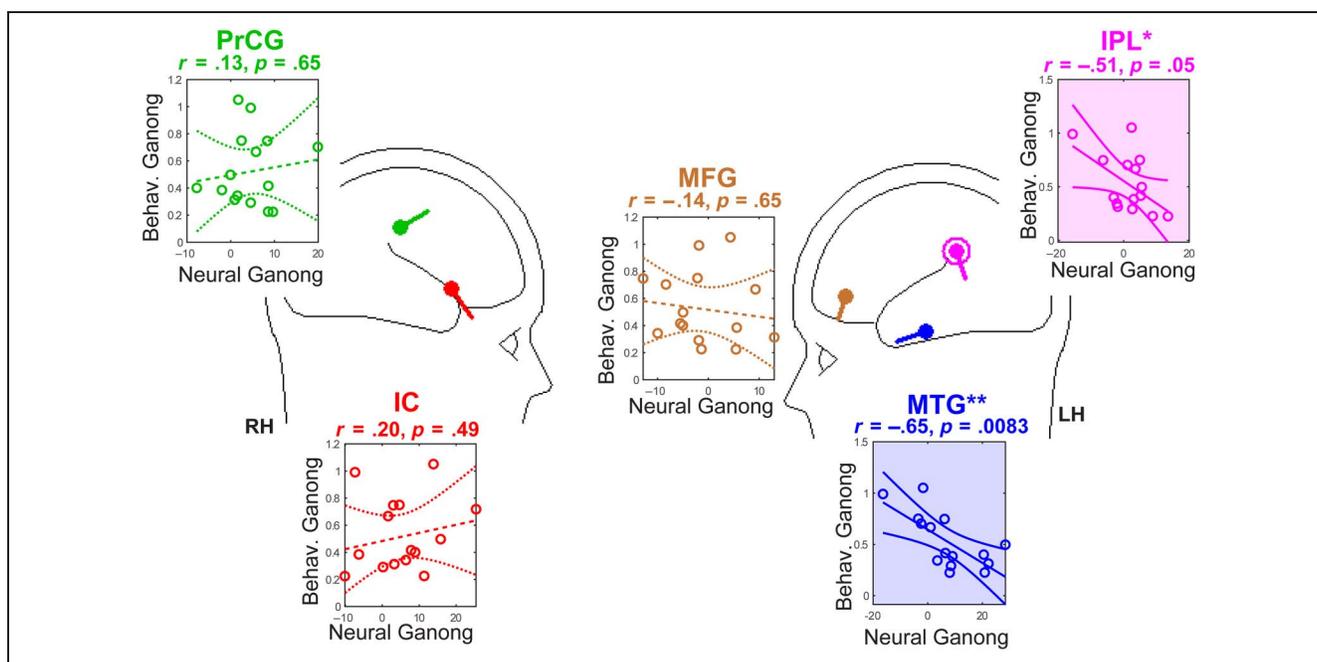


Figure 5. Lexical bias in CP is driven by engagement of MTG and IPL in left hemisphere. Cartoon heads illustrate the location of the dipole sources underlying the neural Ganong effect. Individual scatters show the relation between neural and behavioral Ganong effect measured from each ROI (shading, $p < .05$). Solid regression lines, significant brain–behavior relation; dotted lines, *n.s.* Flanking curved lines reflect 95% CIs. Of the active regions, only left MTG and IPL correspond with listeners’ behavioral bias. LH/RH = left/right hemisphere. * $p < .05$; ** $p < .01$.

compared continuum endpoints that were unequivocally perceived as real words (endpoints perceived as “gift” and “kiss”) with continuum endpoints that were unequivocally perceived as nonwords (endpoints perceived as “giss” and “kift”). These control analyses revealed no significant channel clusters suggesting Ganong differences were not because of the “word status” of the stimuli, per se (Myers & Blumstein, 2008). Similarly, ERP amplitudes for word versus nonword difference waves did not differ from 0 in the same time window that showed Ganong lexical biasing in the experimental conditions, $t(14) = 1.36, p = .20$.

Source analysis of the ERPs exposed neural activations coding lexical bias in CP within five major foci among the auditory-linguistic-motor loop (e.g., Rauschecker & Scott, 2009; Hickok & Poeppel, 2007), including MTG, IPL (proximal to supramarginal gyrus [SMG]), and MFG in left hemisphere, and PrCG and IC in right hemisphere (Figure 4D). For each participant, we extracted the time course of source activity from dipoles seeded at the centroids of these ROIs. We then measured and regressed the peak activation within each ROI (200- to 300-msec analysis window; see Figure 4B)—reflecting the magnitude of “neural Ganong”—against listeners’ behavioral Ganong (i.e., magnitude of perceptual boundary shift; Figure 2C). These brain–behavior correlations revealed strong associations between left MTG and left IPL activity and behavioral bias. The negative association suggests that larger (more positive) change in ERP was associated with smaller magnitude shifts in identification functions. These findings suggest that context-dependent modulations within a restricted temporo-parietal circuit were most inductive to listeners’ susceptibility to lexical influences.

DISCUSSION

By measuring neuroelectric brain activity during rapid speech categorization tasks, our data reveal strong lexical bias in phonetic processing; perception for otherwise identical speech phonemes is attracted toward the direction of words, shifting listeners’ categorical boundary dependent on surrounding speech context. We show a neural analog of lexical biasing emerging within ~200 msec from brain activity localized to a distributed, bilateral temporoparietal network including MTG and IPL. Our findings confirm that when perceiving speech, lexical status rapidly constrains sublexical representations to their category membership within several hundred milliseconds, establishing a direct linguistic influence on early speech processing.

Decoding speech and lexical biasing could be realized via phonetic “feature detectors” (Eimas & Corbit, 1973) that occupy and are differentially sensitive to various segments of the acoustic-linguistic space. Indeed, Ganong-like displacements in perception we observe could occur if linguistic status moves the category boundary toward the most likely lexical candidate. Similarly, nonlinear dynamical models of perception posit that lexical items more strongly activate perceptual “attractor states,” which pull auditory

percepts toward word items (Tuller, Case, Ding, & Kelso, 1994). Under this interpretation, the brain might differentially warp the perceptual space such that even the early acoustic–phonetic analysis of speech is continually anchored to a lexical representation (Liberman, Isenberg, & Rakerd, 1981; Remez, Rubin, Pisoni, & Carrell, 1981).

Considerable debate persists as to whether lexical effects in spoken word recognition result from feedback or feedforward processes (Gow et al., 2008; Myers & Blumstein, 2008; Samuel & Pitt, 2003; Norris et al., 2000; Pitt, 1995). Ganong shifts could occur if lexical knowledge exerts top–down influences to directly affect perceptual states. Under these frameworks, lexical-based modulation of auditory-sensory brain areas (i.e., STG; Myers & Blumstein, 2008; van Linden et al., 2007) could result from top–down input from higher levels associated with word forms (e.g., SMG, MTG). Alternatively, a purely feedforward architecture (Norris et al., 2000) posits that lexical and phonetic outputs combine and interact at later postperceptual stages of processing that are intrinsic to overt perceptual tasks (for illustration of these diametric models, see Figure 1 of Gow et al., 2008; Figure 7: Myers & Blumstein, 2008). In attempts to resolve these conflicting models, Gow et al. (2008) used functional connectivity analyses applied to magnetoencephalography (MEG) data and showed that causal neural signaling directed from left SMG to “lower-level” areas (e.g., STG) modulates sensory representations for speech within a latency of 280–480 msec (Gow et al., 2008). The top–down nature of their effects strongly favored a feedback, perceptual account of the Ganong whereby lexical representations influence the earlier encoding of sublexical speech features (e.g., Noe & Fischer-Baum, 2020; Myers & Blumstein, 2008; van Linden et al., 2007).

Our EEG findings closely agree with MEG data by demonstrating a neural analog of Ganong biasing that unfolds early in the chronometry of speech perception. We observed lexical modulation of speech ERPs beginning ~200 msec after sound onset and no later than 300 msec. The early time window of these effects aligns roughly with the P2 wave of the auditory ERPs, a component that is highly sensitive to perceptual object formation, category structure (Bidelman et al., 2013, 2020; Bidelman & Walker, 2017; Liebenthal et al., 2010), and context effects in speech identification (Bidelman & Lee, 2015).⁴ Two recent EEG studies using Ganong (Noe & Fischer-Baum, 2020) and cross-modal priming (Getz & Toscano, 2019) paradigms suggest even earlier lexical effects in the time-frame of the N1 (75–175 msec; Noe & Fischer-Baum, 2020; Getz & Toscano, 2019). Noe and Fischer-Baum (2020), for example, concluded the early nature of their lexical response at N1 is unlikely to be modulated by top–down effects. Discrepancies between studies as to the time course of lexical effects is unclear but might be attributable to methodological differences.⁵ Categorical effects at N1 have been equivocal in the literature (cf. Noe & Fischer-Baum, 2020; Getz & Toscano, 2019; Bidelman et al., 2013; Toscano, McMurray, Dennhardt, & Luck, 2010;

Sharma & Dorman, 1999). Moreover, previous studies have not adjudicated the underlying sources that contribute to apparent scalp N1 effects. This is important as the N1 wave is composed of sources beyond the supratemporal plane including frontal lobes and IPL (Picton et al., 1999; Woods, 1995; Knight, Hillyard, Woods, & Neville, 1980), areas highly sensitive to lexical influences. Although our data support notions for an early time course of lexical effects (Noe & Fischer-Baum, 2020; Toscano, Anderson, Fabiani, Gratton, & Garnsey, 2018), they also suggest more parallel/iterative influences on perception.

Our data are more consistent with previous source-level MEG findings that demonstrate Ganong-related modulations around 220 msec (Gow et al., 2008). Our source analysis uncovered a Ganong neural circuit spanning five nodes including MTG, IPL, and MFG in the left hemisphere and PrCG, IC in the right hemisphere. The engagement of frontal brain areas (MFG, IC) is consistent with the notion that lexical effects partly evoke postperceptual, executive processes (Norris et al., 2000). The involvement of IC is perhaps also expected in light of prior imaging work; bilateral inferior frontal activation is particularly evident for speech contrasts that are acoustically ambiguous (Feng, Gan, Wan, Wong, & Chandrasekaran, 2018; Bidelman & Dexter, 2015; Guediche, Salvata, & Blumstein, 2013) and under conditions of increased lexical uncertainty (Bidelman & Walker, 2019; Luthra, Guediche, Blumstein, & Myers, 2019) that place higher demands on attention (Bouton et al., 2018). Indeed, resolving phoneme ambiguity (as in the Ganong) may be one of the first processes to come on-line before the decoding of specific lexical features (Gwilliams, Linzen, Poeppel, & Marantz, 2018). This may account for the early time course of our neural effects.

Notable among the Ganong circuit were nodes in left SMG and MTG. Critically, these regions were the only two areas associated with behavior illustrating their important role in the lexical effect. MTG forms a major component of the ventral speech-language pathway that performs sound-to-meaning inference and acts as a lexical interface linking phonological and semantic information (Hickok & Poeppel, 2004, 2007). MTG is also associated with accessing word meaning (Acheson & Hagoort, 2013), a likely operation in our Ganong task when ambiguous phonemes are perceptually (re)interpreted as words. Relatedly, left IPL and adjacent SMG are strongly recruited during auditory phoneme sound categorization (Luthra, Correia, Kleinschmidt, Mesite, & Myers, in press; Desai, Liebenthal, Waldron, & Binder, 2008; Gow et al., 2008), suggesting their role in phonological coding (Sliwiska, Khadilkar, Campbell-Ratcliffe, Quevenco, & Devlin, 2012). Parietal engagement is especially prominent when speech items are more perceptually confusable (Feng et al., 2018) or require added lexical readout as in Ganong paradigms (Oberfeld & Klöckner-Nowotny, 2016) and may serve as the sensory-motor interface for speech (Hickok, Okada, & Serences, 2009; Hickok & Poeppel, 2000).⁶ Moreover, using machine learning to decode full brain EEG, we have

recently shown that left SMG and related outputs from parietal cortex are among the most salient brain areas that code for category decisions (Al-Fahad, Yeasin, & Bidelman, 2020; Mahmud et al., 2020). Similar results were obtained in a multivariate pattern decoding analysis of Luthra et al. (in press), who showed left parietal (SMG) and right temporal (MTG) regions were among the most informative for describing moment-to-moment variability in categorization. In addition, the link between MTG and PrCG implied in our data points to a pathway between the neural substrates that map sounds to meaning and sensorimotor regions that execute motor commands (Al-Fahad et al., 2020; Du, Buchsbaum, Grady, & Alain, 2014). Still, the early time course of these neural effects (~250 msec) occurs well before listeners' behavioral RTs (cf. Figure 2B vs. Figure 4), suggesting these mechanisms operate at an early (pre)perceptual level. These findings lead us to infer that rapid (200–300 msec) context-dependent modulations within a restricted temporo-parietal circuit are most conducive to describing the degree to which listeners are susceptible to lexical influences during speech labeling.

Notably absent from our Ganong circuit—identified via differences waves—was canonical auditory-linguistic brain regions (e.g., STG). Although somewhat unexpected, these data agree with previous fMRI results using a nearly identical Giss–Kiss continuum (Myers & Blumstein, 2008). Indeed, Myers and Blumstein (2008) reported that, for stimulus comparisons at the boundary of a Giss–Kiss gradient (Tk4, as used here), there was strong IPL but no Ganong-related differences in several brain areas previously shown to be sensitive to phonetic category structure including STG and inferior frontal gyrus; STG activation was, however, observed for the boundary condition in a Gift–Kift continuum, suggesting the extent of cortex sensitive to lexical effects depends on the direction and where along the continuum the effect is quantified.⁷ STG activity is greater when stimuli are maximally shifted from their VOT-matched counterparts (Myers & Blumstein, 2008). Although we observe a measurable Ganong effect, it is possible that stronger STG differentiation would have been observed in our EEG data with more salient lexical biasing stimuli (e.g., vowel sounds which are inherently more category ambiguous; Ganong, 1980). Still, the fact that correlations between neural and behavioral Ganong occurred in areas beyond canonical auditory-sensory cortex (e.g., STG) suggests that high-order, top-down mechanisms drive or at least dominate lexical biasing (Gow et al., 2008) rather than auditory temporal cortex, *per se*. Although they do so rapidly, the engagement of a temporo-parietal circuit outside canonical auditory areas (and negative brain–behavior correlations) further implies our lexical effects might be related to decision, attention, or executive control processes. Indeed, IPL is heavily involved in choice decision making, especially during uncertainty (Vickery & Jiang, 2009). This could explain the strong involvement of this region when classifying ambiguous speech in our task. While we cannot

rule out such explanations, the early latency of neural effects (200–300 msec), which occur several hundred milliseconds before listeners' RT decisions, perhaps argues against a straightforward response-selection account of the data. Alternatively, rather than a binary feedforward or feedback model of the lexical effect (Gow et al., 2008), it is possible the formation of speech categories operates in near parallel within lower-order (sensory) and higher-order (cognitive-control) brain structures (Mahmud et al., 2020; Toscano et al., 2018). Our data are broadly consistent with such notions. Category representations also need not be isomorphic across the brain. Category formation might reflect a cascade of events where speech units are reinforced and further discretized by a recontact of acoustic–phonetic with lexical representations (Mahmud et al., 2020; Myers & Blumstein, 2008).

Our data are best cast in terms of interactive rather than serial frameworks of speech perception as in the TRACE model of spoken word recognition (McClelland & Elman, 1986). As confirmed empirically (Noe & Fischer-Baum, 2020; Lam et al., 2017; Gow et al., 2008; Myers & Blumstein, 2008; Ganong, 1980), these models predict stronger lexical biasing when speech sounds carry ambiguity. Indeed, neural correlates of the Ganong effect were most evident at the midpoint of our speech continua, where word influences exert their strongest effect. The very nature of TRACE is that activation traverses from one level to the next before computations at any one stage are complete (McClelland & Elman, 1986). Indeed, available evidence coupled with present results suggest that word recognition could involve simultaneous activation of both continuous acoustic cues and phonological categories (Toscano et al., 2018). It is also possible that the acoustic–phonetic conversion and postperceptual phonetic decision both localize to the same brain areas (Gow et al., 2008, p.621). Nevertheless, our data show that the acoustic–phonetic encoding of speech is rapidly subject to linguistic influences within several hundred milliseconds. While the early time course implies a stage of perceptual processing, we find that lexical effects occur strongest outside the purview of canonical auditory-linguistic brain areas via a restricted temporoparietal circuit.

Acknowledgments

We thank Dr. Emily Myers for sharing stimulus materials.

Reprint requests should be sent to Gavin M. Bidelman, School of Communication Sciences & Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN, 38152, or via e-mail: gmbdlman@memphis.edu.

Funding Information

This work was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award number R01DC016267 (G. M. B.).

Diversity in Citation Practices

A retrospective analysis of the citations in every article published in this journal from 2010 to 2020 has revealed a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were $M(\text{an})/M = .408$, $W(\text{oman})/M = .335$, $M/W = .108$, and $W/W = .149$, the comparable proportions for the articles that these authorship teams cited were $M/M = .579$, $W/M = .243$, $M/W = .102$, and $W/W = .076$ (Fulvio et al., *JoCN*, 33:1, pp. 3–7). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

Notes

1. EEG was not recorded from one participant due to a technical error resulting in a final sample size of $n = 15$ for the neural data (behavioral data were unaffected).
2. Our task blocks stimuli by continuum. One concern is that blocking might set up an expectation such that, upon hearing the initial stop consonant, listeners already have a category structure in mind, biasing their response toward the word end of the continuum. Thus, listeners might “preload” their response, waiting to hear the completion of the word before interpreting the onset consonant as a “g” vs. “k.” Our use of token randomization within each continuum helps prevent such expectancies. Lexical effects are also still observable when stimuli are fully randomized within and across contexts (Ganong, 1980). Moreover, the early time course of our neural Ganong effects (see Figure 4) suggests the brain is already making predictions on lexical status prior to the completion of word endings. Similarly, if listeners know they are in a “gift-kift” block, for example, they may shift their phonetic category boundary more globally such that processing the end of the word (–ift) is no longer necessary. However, one piece of evidence that such global biasing did not occur is that RT speeds were similar for word versus nonword Tk1/Tk8 endpoint tokens (see Figure 2B). Global biasing would be expected to improve decision speeds for tokens heard in a word context.
3. The sign of the difference waveform crucially depends on the order of subtraction (much like an MMN), rendering the direction of the wave somewhat arbitrary. Consequently, we favor an interpretation impartial to direction that implies that because the *change/difference* in response magnitude varies across continua, it is *differential neural activity* that codes the lexical effect.
4. Whether the 200- to 300-msec modulation functionally reflects a late of P2 or early P3 response is unknown. P3 would be more expected in oddball-type paradigms (not used here) but is observable in speech identification tasks although later in time (> 300–400 msec; Bidelman & Alain, 2015; Toscano et al., 2010). A similar “post-P2” wave (180–320 msec) has been reported during speech categorization (Bidelman & Alain, 2015; Bidelman et al., 2013), which varied with perceptual (rather than acoustic) classification. This response could represent an integration or reconciliation of the input with a phonetic memory template (Bidelman, Bush, & Boudreaux, 2020; Bidelman & Alain, 2015) and/or attentional reorienting during stimulus evaluation (Knight, Scabini, Woods, & Clayworth, 1989). In support of this functional interpretation, the modulation is observed when classifying speech under higher levels of uncertainty, for example, when identifying speech in noise (Bidelman et al., 2020).

5. Both N1 studies (Noe & Fischer-Baum, 2020; Getz & Toscano, 2019) used average mastoid reference recordings, which can inflate and bias neural effects to frontal electrodes (Yao et al., 2005) where their ERPs were quantified. Here, we used average reference data (and source imaging), which provides a less biased and unmixed view of neural activity. Another notable difference in Noe and Fischer-Baum (2020) is their use of single trials ($n = 38,491$ observations) in the statistical analysis to detect lexical effects at N1. Although independence assumptions of using such large quantities of correlated trial-wise EEG might be debatable, such analyses might be more sensitive to detecting earlier lexical effects than the subject-wise approach used here.
6. The basis of the negative correlation between “neural” and behavioral Ganong is not entirely clear; positive associations are more easily hypothesized. Speculatively, the negative relation could be related to a lexical ambiguity interpretation. Thus, the negative correlation we find between IPL and MTG and behavioral Ganong shifts (Figure 5) might occur if larger degrees of ambiguity between speech sounds (evoking larger ERP differences waves) reduces lexical certainty. This would tend to reduce the magnitude of the perceptual lexical effect as seen behaviorally.
7. Myers and Blumstein (2008, p. 283) reported strong clusters of lexically sensitive cortex in canonical auditory areas including STG for boundary stimulus comparisons in their Gift–Kift continuum. In that study, the boundary condition was defined based on the perceptual location within each continuum (i.e., Tk 5 for Gift–Kift; Tk 4 for Giss–Kiss). Here, we compared activation patterns solely at the physically identical Tk 4 stimulus (where perceptual lexical effects were maximal; Figure 2).

REFERENCES

- Acheson, D. J., & Hagoort, P. (2013). Stimulating the brain's language network: Syntactic ambiguity resolution after TMS to the inferior frontal gyrus and middle temporal gyrus. *Journal of Cognitive Neuroscience*, *25*, 1664–1677. **DOI:** https://doi.org/10.1162/jocn_a.00430, **PMID:** 23767923
- Al-Fahad, R., Yeasin, M., & Bidelman, G. M. (2020). Decoding of single-trial EEG reveals unique states of functional brain connectivity that drive rapid speech categorization decisions. *Journal of Neural Engineering*, *17*, 016045. **DOI:** <https://doi.org/10.1088/1741-2552/ab6040>, **PMID:** 31822643
- 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*, 1–11. **DOI:** <https://doi.org/10.1038/srep40790>, **PMID:** 28102300, **PMCID:** PMC5244401
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, *57*, 289–300. **DOI:** <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Best, R. M., & Goldstone, R. L. (2019). Bias to (and away from) the extreme: Comparing two models of categorical perception effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 1166–1176. **DOI:** <https://doi.org/10.1037/xlm0000609>, **PMID:** 30024260
- Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *NeuroImage*, *175*, 56–69. **DOI:** <https://doi.org/10.1016/j.neuroimage.2018.03.060>, **PMID:** 29604459
- Bidelman, G. M. & Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *Journal of Neuroscience*, *35*, 1240–1249. **DOI:** <https://doi.org/10.1523/JNEUROSCI.3292-14.2015>, **PMID:** 25609638, **PMCID:** PMC6605547
- Bidelman, G. M., Bush, L. C., & Boudreaux, A. M. (2020). Effects of noise on the behavioral and neural categorization of speech. *Frontiers in Neuroscience*, *14*, 1–13. **DOI:** <https://doi.org/10.3389/fnins.2020.00153>, **PMID:** 32180700, **PMCID:** PMC7057933
- 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. **DOI:** <https://doi.org/10.1016/j.heares.2018.05.018>, **PMID:** 29871826, **PMCID:** PMC6105463
- Bidelman, G. M., & Dexter, L. (2015). Bilinguals at the “cocktail party”: Dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners’ speech-in-noise recognition deficits. *Brain and Language*, *143*, 32–41. **DOI:** <https://doi.org/10.1016/j.bandl.2015.02.002>, **PMID:** 25747886
- Bidelman, G. M., & Lee, C.-C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *NeuroImage*, *120*, 191–200. **DOI:** <https://doi.org/10.1016/j.neuroimage.2015.06.087>, **PMID:** 26146197
- Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, *79*, 201–212. **DOI:** <https://doi.org/10.1016/j.neuroimage.2013.04.093>, **PMID:** 23648960
- Bidelman, G. M., & Walker, B. (2017). Attentional modulation and domain specificity underlying the neural organization of auditory categorical perception. *European Journal of Neuroscience*, *45*, 690–699. **DOI:** <https://doi.org/10.1111/ejn.13526>, **PMID:** 28112440
- Bidelman, G. M., & Walker, B. S. (2019). Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network. *NeuroImage*, *201*, 1–10. **DOI:** <https://doi.org/10.1016/j.neuroimage.2019.116022>, **PMID:** 31310863, **PMCID:** PMC6765438
- Bouton, S., Chambon, V., Tyrand, R., Guggisberg, A. G., Seeck, M., Karkar, S., et al. (2018). Focal versus distributed temporal cortex activity for speech sound category assignment. *Proceedings of the National Academy of Sciences, U.S.A.*, *115*, E1299–E1308. **DOI:** <https://doi.org/10.1073/pnas.1714279115>, **PMID:** 29363598, **PMCID:** PMC5819402
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum Associates.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. **DOI:** <https://doi.org/10.1016/j.jneumeth.2003.10.009>, **PMID:** 15102499
- Desai, R., Liebenthal, E., Waldron, E., & Binder, J. R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *Journal of Cognitive Neuroscience*, *20*, 1174–1188. **DOI:** <https://doi.org/10.1162/jocn.2008.20081>, **PMID:** 18284339, **PMCID:** PMC3350814
- Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences, U.S.A.*, *111*, 1–6. **DOI:** <https://doi.org/10.1073/pnas.1318738111>, **PMID:** 24778251, **PMCID:** PMC4024897
- Dunlap, W., Cortina, J., Vaslow, J., & Burke, M. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*, 170–177. **DOI:** <https://doi.org/10.1037/1082-989X.1.2.170>
- Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*, 99–109. **DOI:** [https://doi.org/10.1016/0010-0285\(73\)90006-6](https://doi.org/10.1016/0010-0285(73)90006-6)

- Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27, 143–165. **DOI:** [https://doi.org/10.1016/0749-596X\(88\)90071-X](https://doi.org/10.1016/0749-596X(88)90071-X)
- Feng, G., Gan, S., Wan, S., Wong, P. C. M., & Chandrasekaran, B. (2018). Task-general and acoustic-invariant neural representation of speech categories in the human brain. *Cerebral Cortex*, 28, 3241–3254. **DOI:** <https://doi.org/10.1093/cercor/bhx195>, **PMID:** 28968658, **PMCID:** PMC6454529
- Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 526. **DOI:** <https://doi.org/10.1037/0096-1523.10.4.526>
- Francis, A. L., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *Journal of the Acoustical Society of America*, 114, 1611–1621. **DOI:** <https://doi.org/10.1121/1.1603231>, **PMID:** 14514214
- Ganong, W. F., III. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110–125. **DOI:** <https://doi.org/10.1037/0096-1523.6.1.110>
- Ganong, W. F., & Zatorre, R. J. (1980). Measuring phoneme boundaries four ways. *Journal of the Acoustical Society of America*, 68, 431–439. **DOI:** <https://doi.org/10.1121/1.384754>
- Getz, L. M., & Toscano, J. C. (2019). Electrophysiological evidence for top-down lexical influences on early speech perception. *Psychological Science*, 30, 830–841. **DOI:** <https://doi.org/10.1177/0956797619841813>, **PMID:** 31018103
- Goldstone, R. L., & Hendrickson, A. T. 2010. Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 69–78. **DOI:** <https://doi.org/10.1002/wcs.26>, **PMID:** 26272840
- Goldstone, R. L., Steyvers, M., Spencer-Smith, J., & Kersten, A. (2000). Interactions between perceptual and conceptual learning. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 191–228). Mahwah, NJ: Lawrence Erlbaum.
- Gow, D. W., Jr., Segawa, J. A., Ahlfors, S. P., & Lin, F.-H. (2008). Lexical influences on speech perception: a Granger causality analysis of MEG and EEG source estimates. *NeuroImage*, 43, 614–623. **DOI:** <https://doi.org/10.1016/j.neuroimage.2008.07.027>, **PMID:** 18703146, **PMCID:** PMC2585985
- Guediche, S., Salvata, C., & Blumstein, S. E. (2013). Temporal cortex reflects effects of sentence context on phonetic processing. *Journal of Cognitive Neuroscience*, 25, 706–718. **DOI:** https://doi.org/10.1162/jocn_a_00351, **PMID:** 23281778, **PMCID:** PMC3612392
- Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology*, 28, 240–244. **DOI:** <https://doi.org/10.1111/j.1469-8986.1991.tb00417.x>, **PMID:** 1946890
- Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In spoken word recognition, the future predicts the past. *Journal of Neuroscience*, 38, 7585–7599. **DOI:** <https://doi.org/10.1523/JNEUROSCI.0065-18.2018>, **PMID:** 30012695, **PMCID:** PMC6113903
- Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*, 101, 2725–2732. **DOI:** <https://doi.org/10.1152/jn.91099.2008>, **PMID:** 19225172
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4, 131–138. **DOI:** [https://doi.org/10.1016/S1364-6613\(00\)01463-7](https://doi.org/10.1016/S1364-6613(00)01463-7)
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99. **DOI:** <https://doi.org/10.1016/j.cognition.2003.10.011>, **PMID:** 15037127
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402. **DOI:** <https://doi.org/10.1038/nrn2113>, **PMID:** 17431404
- Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, 72, 1218–1227. **DOI:** <https://doi.org/10.3758/APP.72.5.1218>, **PMID:** 20601702, **PMCID:** PMC2921848
- Jordanov, T., Hoehstetter, K., Berg, P., Paul-Jordanov, I., & Scherg, M. (2014). CLARA: Classical LORETA analysis recursively applied. OHBM 2014.
- Knight, R. T., Hillyard, S. A., Woods, D. L., & Neville, H. J. (1980). The effects of frontal and temporal-parietal lesions on the auditory evoked potential in man. *Clinical Neurophysiology*, 50, 112–124. **DOI:** [https://doi.org/10.1016/0013-4694\(80\)90328-4](https://doi.org/10.1016/0013-4694(80)90328-4)
- Knight, R. T., Scabini, D., Woods, D. L., & Clayworth, C. C. (1989). Contributions of temporal-parietal junction to the human auditory P3. *Brain Research*, 502, 109–116. **DOI:** [https://doi.org/10.1016/0006-8993\(89\)90466-6](https://doi.org/10.1016/0006-8993(89)90466-6)
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606–608. **DOI:** <https://doi.org/10.1126/science.1736364>, **PMID:** 1736364
- Lam, B. P. W., Xie, Z., Tessmer, R., & Chandrasekaran, B. (2017). The downside of greater lexical influences: Selectively poorer speech perception in noise. *Journal of Speech, Language, and Hearing Research*, 60, 1662–1673. **DOI:** https://doi.org/10.1044/2017_JSLHR-H-16-0133, **PMID:** 28586824, **PMCID:** PMC5544416
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461. **DOI:** <https://doi.org/10.1037/h0020279>, **PMID:** 4170865
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phonemic boundaries. *Journal of Experimental Psychology*, 54, 358–368. **DOI:** <https://doi.org/10.1037/h0044417>, **PMID:** 13481283
- Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception and Psychophysics*, 30, 133–143. **DOI:** <https://doi.org/10.3758/BF03204471>, **PMID:** 7301513
- Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cerebral Cortex*, 20, 2958–2970. **DOI:** <https://doi.org/10.1093/cercor/bhq045>, **PMID:** 20382643, **PMCID:** PMC2978244
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, 94, 1242–1255. **DOI:** <https://doi.org/10.1121/1.408177>, **PMID:** 8408964, **PMCID:** PMC3509365
- Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 732–753. **DOI:** <https://doi.org/10.1037/0278-7393.24.3.732>
- Lotto, A. J., & Holt, L. L. (2016). Speech perception: The view from the auditory system. In *Neurobiology of language* (pp. 185–194). Elsevier. **DOI:** <https://doi.org/10.1016/B978-0-12-407794-2.00016-X>
- Luthra, S., Guediche, S., Blumstein, S. E., & Myers, E. B. (2019). Neural substrates of subphonemic variation and lexical competition in spoken word recognition. *Language*,

- Cognition and Neuroscience*, 34, 151–169. **DOI:** <https://doi.org/10.1080/23273798.2018.1531140>, **PMID:** 31106225, **PMCID:** PMC6516505
- Luthra, S. C., Correia, J. M., Kleinschmidt, D. F., Mesite, L., & Myers, E. B. (in press). Lexical information guides retuning of neural patterns in perceptual learning for speech. *Journal of Cognitive Neuroscience*, 1–12. **DOI:** https://doi.org/10.1162/jocn_a_01612, **PMID:** 32662731
- Mahmud, S., Yeasin, M., Bidelman, G. M. (2020). Data-driven machine learning models for decoding speech categorization from evoked brain responses. *bioRxiv* [preprint]. **DOI:** <https://doi.org/10.1101/2020.08.03.234997>
- Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences, U.S.A.*, 115, 13129–13134. **DOI:** <https://doi.org/10.1073/pnas.1811793115>, **PMID:** 30509989, **PMCID:** PMC6304957
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86. **DOI:** [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- McMurray, B., Dennhardt, J. L., & Struck-Marcell, A. (2008). Context effects on musical chord categorization: Different forms of top-down feedback in speech and music? *Cognitive Science*, 32, 893–920. **DOI:** <https://doi.org/10.1080/03640210802222021>, **PMID:** 21490878, **PMCID:** PMC3073321
- Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology*, 115, 2195–2222. **DOI:** <https://doi.org/10.1016/j.clinph.2004.06.001>, **PMID:** 15351361
- Myers, E. B., & Blumstein, S. E. (2008). The neural bases of the lexical effect: An fMRI investigation. *Cerebral Cortex*, 18, 278–288. **DOI:** <https://doi.org/10.1093/cercor/bhm053>, **PMID:** 17504782
- Noe, C., & Fischer-Baum, S. (2020). Early lexical influences on sublexical processing in speech perception: Evidence from electrophysiology. *Cognition*, 197, 104162. **DOI:** <https://doi.org/10.1016/j.cognition.2019.104162>, **PMID:** 31901875
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299–325; discussion 325–270. **DOI:** <https://doi.org/10.1017/S0140525X00003241>, **PMID:** 11301575
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238. **DOI:** [https://doi.org/10.1016/S0010-0285\(03\)00006-9](https://doi.org/10.1016/S0010-0285(03)00006-9)
- Oberfeld, D., & Klöckner-Nowotny, F. (2016). Individual differences in selective attention predict speech identification at a cocktail party. *eLife*, 5. **DOI:** <https://doi.org/10.7554/eLife.16747>, **PMID:** 27580272, **PMCID:** PMC5441891
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113. **DOI:** [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, 112, 713–719. **DOI:** [https://doi.org/10.1016/S1388-2457\(00\)00527-7](https://doi.org/10.1016/S1388-2457(00)00527-7)
- Pascual-Marqui, R. D., Esslen, M., Kochi, K., & Lehmann, D. (2002). Functional imaging with low-resolution brain electromagnetic tomography (LORETA): A review. *Methods and Findings in Experimental and Clinical Pharmacology*, 24 Suppl. C, 91–95.
- Phillips, C. (2001). Levels of representation in the electrophysiology of speech perception. *Cognitive Science*, 25, 711–731. **DOI:** https://doi.org/10.1207/s15516709cog2505_5
- Picton, T. W., Alain, C., Woods, D. L., John, M. S., Scherg, M., Valdes-Sosa, P., et al. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neuro-Otology*, 4, 64–79. **DOI:** <https://doi.org/10.1159/000013823>, **PMID:** 9892757
- Picton, T. W., van Roon, P., Armiljo, M. L., Berg, P., Ille, N., & Scherg, M. (2000). The correction of ocular artifacts: A topographic perspective. *Clinical Neurophysiology*, 111, 53–65. **DOI:** [https://doi.org/10.1016/S1388-2457\(99\)00227-8](https://doi.org/10.1016/S1388-2457(99)00227-8)
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 13, 253–260. **DOI:** <https://doi.org/10.3758/BF03214136>, **PMID:** 23226880, **PMCID:** PMC3515632
- Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory and Cognition*, 3, 7–18. **DOI:** <https://doi.org/10.3758/BF03198202>, **PMID:** 24203819, **PMCID:** PMC3831373
- Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, 25, 21–52. **DOI:** [https://doi.org/10.1016/0010-0277\(87\)90003-5](https://doi.org/10.1016/0010-0277(87)90003-5)
- Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics*, 15, 285–290. **DOI:** <https://doi.org/10.3758/BF03213946>, **PMID:** 23226881, **PMCID:** PMC3515635
- Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1037–1052. **DOI:** <https://doi.org/10.1037/0278-7393.21.4.1037>, **PMID:** 7673866
- Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 699–725. **DOI:** <https://doi.org/10.1037/0096-1523.19.4.699>
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12, 718–724. **DOI:** <https://doi.org/10.1038/nn.2331>, **PMID:** 19471271, **PMCID:** PMC2846110
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212, 947–949. **DOI:** <https://doi.org/10.1126/science.7233191>, **PMID:** 7233191
- Repp, B. H., & Liberman, A. M. (1987). Phonetic category boundaries are flexible. In S. R. Harnad (Ed.), *Categorical perception: the groundwork of cognition* (pp. 89–112). New York: Cambridge University Press.
- Richards, J. E., Sanchez, C., Phillips-Meek, M., & Xie, W. (2016). A database of age-appropriate average MRI templates. *NeuroImage*, 124, 1254–1259. **DOI:** <https://doi.org/10.1016/j.neuroimage.2015.04.055>, **PMID:** 25941089, **PMCID:** PMC4630162
- Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, 48. **DOI:** [https://doi.org/10.1016/S0749-596X\(02\)00514-4](https://doi.org/10.1016/S0749-596X(02)00514-4)
- Scherg, M., Berg, P., Nakasato, N., & Beniczky, S. (2019). Taking the EEG back into the brain: The power of multiple discrete sources. *Frontiers in Neurology*, 10. **DOI:** <https://doi.org/10.3389/fneur.2019.00855>, **PMID:** 31481921, **PMCID:** PMC6710389
- Scherg, M., Ille, N., Bornfleth, H., & Berg, P. (2002). Advanced tools for digital EEG review: Virtual source montages, whole-head mapping, correlation, and phase analysis. *Journal of Clinical Neurophysiology*, 19, 91–112. **DOI:** <https://doi.org/10.1097/00004691-200203000-00001>, **PMID:** 11997721
- Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America*, 106, 1078–1083. **DOI:** <https://doi.org/10.1121/1.428048>, **PMID:** 10462812

- Sliwinska, M. W., Khadilkar, M., Campbell-Ratcliffe, J., Quevenco, F., & Devlin, J. T. (2012). Early and sustained supramarginal gyrus contributions to phonological processing. *Frontiers in Psychology, 3*. DOI: <https://doi.org/10.3389/fpsyg.2012.00161>, PMID: 22654779, PMCID: PMC3361019
- Toscano, J. C., Anderson, N. D., Fabiani, M., Gratton, G., & Garnsey, S. M. (2018). The time-course of cortical responses to speech revealed by fast optical imaging. *Brain and Language, 184*, 32–42. DOI: <https://doi.org/10.1016/j.bandl.2018.06.006>, PMID: 29960165, PMCID: PMC6102048
- Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science, 21*, 1532–1540. DOI: <https://doi.org/10.1177/0956797610384142>, PMID: 20935168, PMCID: PMC3523688
- Tuller, B., Case, P., Ding, M., & Kelso, J. A. S. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 3–16. DOI: <https://doi.org/10.1037/0096-1523.20.1.3>
- van Linden, S., Stekelenburg, J. J., Tuomainen, J., & Vroomen, J. (2007). Lexical effects on auditory speech perception: An electrophysiological study. *Neuroscience Letters, 420*, 49–52. DOI: <https://doi.org/10.1016/j.neulet.2007.04.006>, PMID: 17475404
- Vickery, T. J. & Jiang, Y. V. (2009). Inferior parietal lobule supports decision making under uncertainty in humans. *Cerebral Cortex, 19*, 916–925. DOI: <https://doi.org/10.1093/cercor/bhn140>, PMID: 18728197
- Woods, D. L. (1995). The component structure of the N1 wave of the human auditory evoked potential. *Electroencephalography and Clinical Neurophysiology, 44*, 102–109.
- Yao, D., Wang, L., Oostenveld, R., Nielsen, K. D., Arendt-Nielsen, L., & Chen, A. C. (2005). A comparative study of different references for EEG spectral mapping: the issue of the neutral reference and the use of the infinity reference. *Physiological Measurement, 26*, 173–184. DOI: <https://doi.org/10.1088/0967-3334/26/3/003>, PMID: 15798293