

## Short Communication

## Induced neural beta oscillations predict categorical speech perception abilities



Gavin M. Bidelman\*

Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA  
 School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA

## ARTICLE INFO

Article history:  
 Accepted 9 November 2014

Keywords:  
 Beta activity  
 EEG  
 Categorical speech perception  
 Gamma activity  
 Induced brain responses  
 Neural oscillations  
 Time–frequency analysis

## ABSTRACT

Neural oscillations have been linked to various perceptual and cognitive brain operations. Here, we examined the role of these induced brain responses in categorical speech perception (CP), a phenomenon in which similar features are mapped to discrete, common identities despite their equidistant/continuous physical spacing. We recorded neuroelectric activity while participants rapidly classified sounds along a vowel continuum (/u/ to /a/). Time–frequency analyses applied to the EEG revealed distinct temporal dynamics in induced (non-phase locked) oscillations; increased  $\beta$  (15–30 Hz) coded prototypical vowel sounds carrying well-defined phonetic categories whereas increased  $\gamma$  (50–70 Hz) accompanied ambiguous tokens near the categorical boundary. Notably, changes in  $\beta$  activity were strongly correlated with the slope of listeners' psychometric identification functions, a measure of the “steepness” of their categorical percept. Our findings demonstrate that in addition to previously observed evoked (phase-locked) correlates of CP, induced brain activity in the  $\beta$ -band codes the ambiguity and strength of categorical speech percepts.

© 2014 Elsevier Inc. All rights reserved.

## 1. Introduction

Speech perception requires that the human brain warp continuous variations in stimulus acoustics into discrete perceptual units, a conversion exemplified in the phenomenon of categorical perception (CP). CP is particularly salient in the context of speech listening whereby gradually morphed sounds along a large acoustic continuum are heard as belonging to one of only a few discrete phonetic classes (Bidelman, Moreno, & Alain, 2013; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Pisoni, 1973; Pisoni & Luce, 1987). At extreme ends of the continuum, tokens are perceived as having the same phonetic identity. Despite equal physical spacing between adjacent steps, a sharp change in percept occurs near the midpoint of the stimulus continuum, marking the categorical boundary.

Categorical speech boundaries emerge early in life well before the development of the lexical-semantic system (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). The fundamental importance of this “phonetic mode” of listening to speech and language processing is evident by its integral role in speech acquisition (Eimas et al., 1971) and the grapheme-to-phoneme mapping essential

for reading and writing skills (Werker & Tees, 1987). Interestingly, the categorical nature of speech perception can be further modified based on one's native tongue (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Xu, Gandour, & Francis, 2006), normal aging and hearing loss (Bidelman, Villafuerte, Moreno, & Alain, 2014a; Strouse, Ashmead, Ohde, & Grantham, 1998), and the plastic effects of auditory training (Bidelman, Weiss, Moreno, & Alain, 2014b). Thus, while the neural mechanisms underlying CP are potentially innate (Eimas et al., 1971), they are also influenced by experiential factors (Bidelman et al., 2014b). The fact that speech sounds separated by equal acoustic distance are nevertheless perceived in an unequal, categorical fashion suggests that in order to generate adequate speech percepts, the auditory system must transform the continuous acoustic domain of the external world into the discrete perceptual space internalized by a listener (Bidelman et al., 2013; Iverson & Kuhl, 1995; Pisoni & Luce, 1987). Understanding how this ubiquitous acoustic–phonetic transformation is realized within the human brain is among the many broad and widespread interests to understand how sensory features are mapped to higher-order perceptual attributes (Bidelman et al., 2013; Phillips, 2001; Pisoni & Luce, 1987).

The majority of neuroimaging work examining CP has focused on identifying its neuroanatomical locus, i.e., brain areas which show a “categorical neural organization” (e.g., Binder, Liebenthal, Possing, Medler, & Ward, 2004; Chang et al., 2010; Zhang et al.,

\* Address: School of Communication Sciences & Disorders, University of Memphis, 807 Jefferson Ave., Memphis, TN 38105, USA. Fax: +1 (901) 525 1282.

E-mail address: [g.bidelman@memphis.edu](mailto:g.bidelman@memphis.edu)

2011). Studies demonstrate that auditory processing in early superior temporal gyrus supports such a categorical scheme and codes the phonetic (higher-order cognitive code), rather than solely acoustic features of speech (low-level perceptual code). Additionally, categorical neural organization has been observed in inferior frontal gyrus (Du, Buchsbaum, Grady, & Alain, 2014; Myers, Blumstein, Walsh, & Eliassen, 2009) and premotor cortices (Chevillet, Jiang, Rauschecker, & Riesenhuber, 2013; Du et al., 2014), suggesting a diffuse yet expansive code for CP that recruits both early auditory areas as well as brain regions downstream from lemniscal auditory structures. In our recent study, we traced the neural chronometry of CP by comparing the differential magnitude of event-related brain potentials (ERPs) evoked by tokens along a vowel continuum (Bidelman et al., 2013). Robust neural correlates of CP emerged in ERP amplitude within the timeframe of the auditory N1 wave (~100 ms) and were fully manifested by P2 (~150–200 ms). These results converged with human intracranial recordings (Chang et al., 2010) to suggest that the neurophysiological underpinnings of categorical speech perception are present within the first few hundred milliseconds after sound enters the ear. While these studies shed light on early evoked cortical activity underlying the neural encoding of speech, they cannot speak to potential connections between induced (i.e., intrinsic) brain responses and CP.

Human EEG activity can be broadly defined as either phase-locked (“evoked”) or non-phase-locked (“induced”) to experimental events (Pantev, 1995; Pfurtscheller & Lopes da Silva, 1999; Shahin, Picton, & Miller, 2009). The former is traditionally characterized via the ERPs and reflects repeatable, phase-coherent potentials generated in response to a time-locking stimulus. In addition to evoked components, modulations in the ongoing “rhythms” of the electroencephalogram (i.e., “neural oscillations”), while not directly phase-locked to the stimulus, can nevertheless be induced by perceptual or cognitive processes recruited during various tasks. The latter is unobservable in traditional evoked potentials (ERPs) as phase discrepancies (i.e., temporal jitter) between trials causes induced responses to cancel in the time-locked average (Picton, 2010, p. 65). Collectively, ERPs and induced patterns of the EEG provide unique neural signatures which carry different functional relevance to auditory perception and speech-language processing (Foxe & Snyder, 2011; Giraud & Poeppel, 2012; Pantev, 1995; Shahin et al., 2009). Indeed, certain frequency bands of the EEG have been linked to different neurocognitive operations and various levels of temporal ‘granularity’ subserving language processing (e.g., Giraud & Poeppel, 2012; von Stein & Sarnthein, 2000). For example, increased activity in the  $\gamma$ -band (30–120 Hz) is associated with local network synchronization (Giraud & Poeppel, 2012; Haenschel, Baldeweg, Croft, Whittington, & Gruzelier, 2000), the formation of auditory objects (Tallon-Baudry & Bertrand, 1999), and semantic evaluation of speech information (Shahin et al., 2009). In contrast, activity in the  $\beta$ -band (15–30 Hz) is associated with template matching between stimuli and internalized representations stored in auditory/lexical memory (Shahin et al., 2009).  $\beta$  oscillations also play a prominent role in perceptual binding and network interactions across modalities (e.g., Aissani, Martinerie, Yahia-Cherif, Paradis, & Lorenceau, 2014; Brovelli et al., 2004; von Stein & Sarnthein, 2000), processes that are precursors to template matching. Of interest to the current study, are the temporal dynamics of these brain oscillations and which spectral profiles relate to sound identification and the categorical nature of speech perception.

While recording neuroelectric activity, we asked participants to identify speech sounds along a vowel continuum in a speeded classification task (Bidelman et al., 2013, 2014a,b). Listeners heard 200 randomly ordered exemplars of each token and were asked to label them with a binary response as quickly as possible (“u” or “a”).

Analysis and results of the evoked potential components of this dataset, including the modulation of the cortical N1-P2 waves with listeners’ CP are reported in our previous study (Bidelman et al., 2013). New time–frequency analysis of the EEG provided novel insight into the neural correlates of CP and phonetic classification with respect to induced oscillatory brain activity. Under the hypothesis that neural oscillations in the  $\beta$ - and  $\gamma$ -band reflect lexical-semantic processes (Shahin et al., 2009) and the formation of auditory object representations (Tallon-Baudry & Bertrand, 1999), we expected modulations in these rhythms to covary with listeners’ degree of CP for speech and/or how well tokens matched phonetic categories. Specifically, we expected stronger  $\beta$  and/or  $\gamma$ -band responses for unambiguous speech tokens (i.e., those heard with distinct phonetic categories) relative to ambiguous sounds as the former evokes a stronger match to internalized speech templates.

## 2. Results

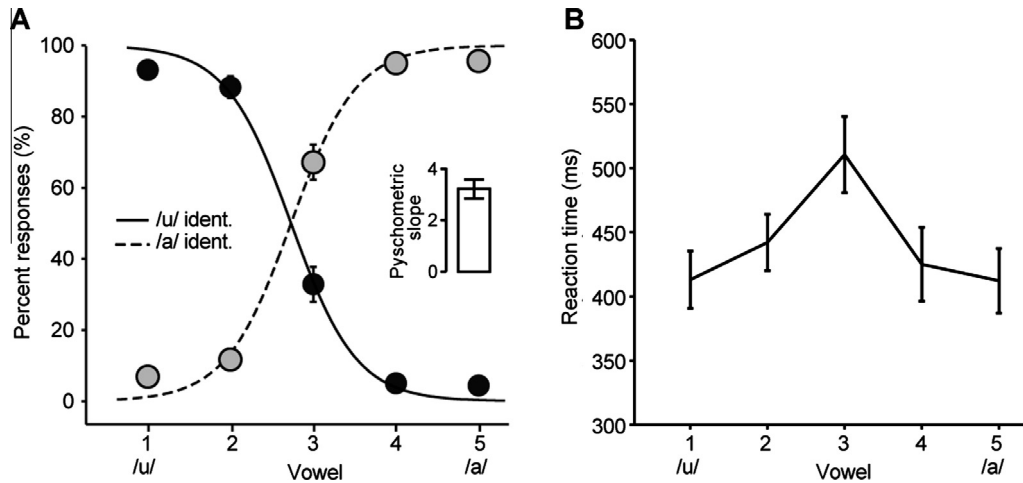
### 2.1. Behavioral data

Behavioral speech identification functions are shown in Fig. 1A. Despite equidistant spacing between stimulus first-formant frequency, tokens near the low end of the vowel continuum (e.g., vws 1–2) were heard as belonging exclusively to the /u/ phonetic category, whereas tokens at the opposite end were heard as /a/ [rmANOVA,  $F_{4,44} = 272.5$ ,  $p < 0.001$ ]. Listeners heard an abrupt perceptual shift in speech identity (/u/ vs. /a/) near the middle of the continuum (vw 3), a hallmark of categorical speech perception (Bidelman et al., 2013, 2014a,b; Chang et al., 2010; May, 1981; Pisoni, 1973). The slope of listeners’ psychometric identification functions ( $\beta_1$  parameter of sigmoidal fits), an index of the “steepness” of the categorical boundary, was  $3.22 \pm 0.37$  (Fig. 1A, inset).

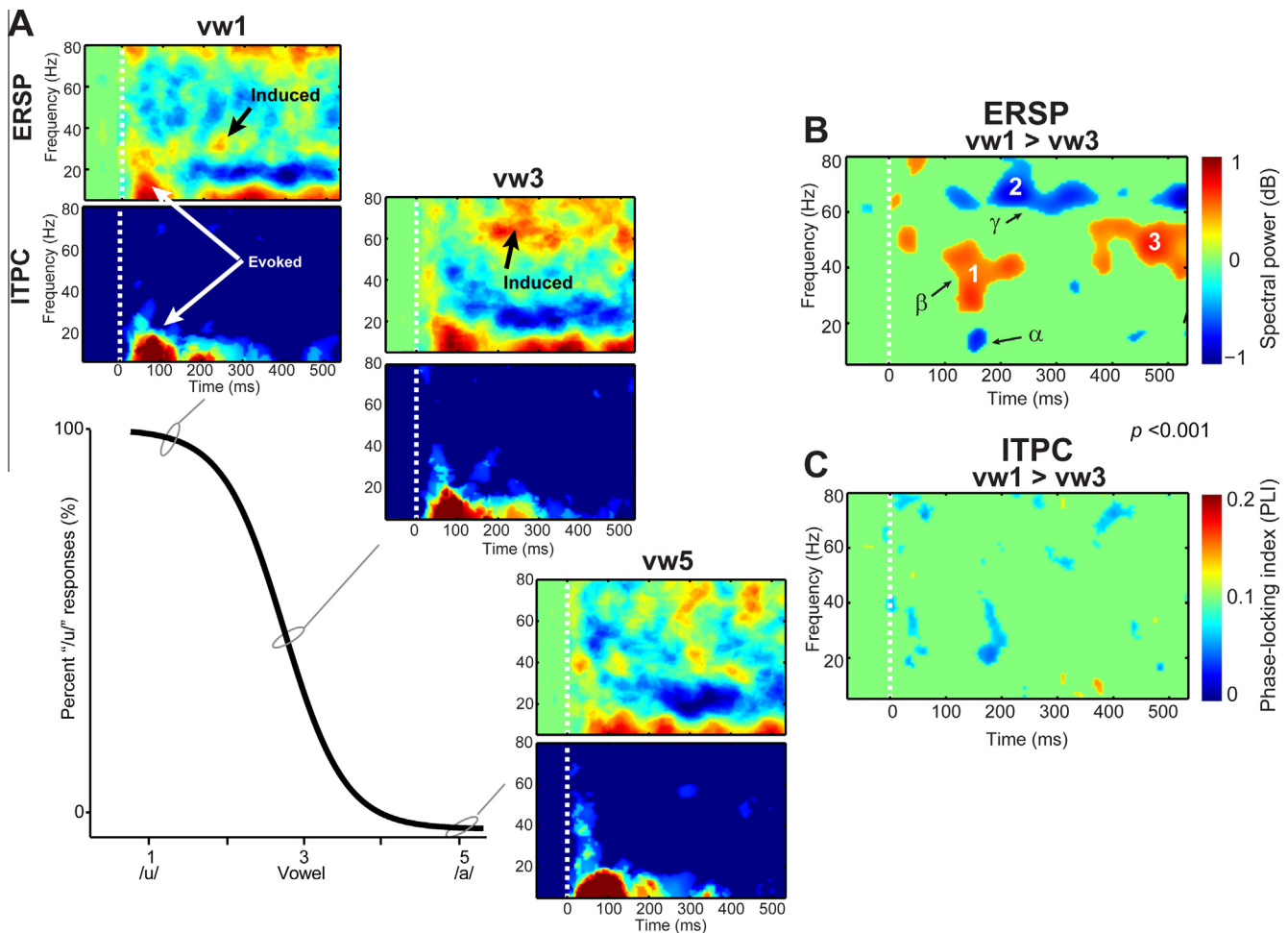
Speech labeling speeds (i.e., reaction times, RTs) are shown in Fig. 1B. Listeners were slower [ $F_{4,44} = 23.62$ ,  $P < 0.001$ ] when classifying tokens surrounding the categorical boundary compared to those within either category ( $P < 0.001$ , Bonferroni corrected). This suggests that tokens near the perceptual boundary (vw3) were more ambiguous and difficult to classify than prototypical tokens at the ends of the vowel continuum (vw1 and vw5).

### 2.2. Neural oscillatory responses to CP stimuli

Event-related Spectral Perturbation (ERSP) and Inter-Trial Phase Coherence (ITPC) spectrographic maps are shown for the two prototypical (vw1, vw5) and ambiguous (vw3) tokens along the speech continuum in Fig. 2. ERSPs reflect both evoked and induced changes in EEG power relative to the pre-stimulus baseline; ITPCs reflect phase-locked neural activity consistent across stimulus trials and are akin to the power spectrum of the ERP. Oscillations that occur in the ERSP but not the ITPC are indicative of “induced” neural processing not directly phase-locked to the auditory presentation, but nevertheless generated in response to the speech stimulus. Both ERSPs and ITPCs show strong modulations in the  $\alpha$ -band (8–13 Hz) circumscribed to the first 100–200 ms of post-stimulus activity. However, the presence of these oscillations in both the ERSP and ITPC indicate this activity is evoked, i.e., phase-locked to the eliciting speech sounds and likely a component(s) of the ERPs. Vowels at the continuum’s endpoints (vw1 and vw5) elicited similar ERSP/ITPC responses, likely given the symmetric nature of categorical perception (e.g., Fig. 1). Hence, a majority of subsequent analyses focused on comparisons between vw1 and vw3, representative of a prototypical token at the continuum’s endpoint and one nearest the perceptual boundary, respectively.



**Fig. 1.** Psychometric identification functions for speech sound classification. (A) Despite the continuous acoustic change between vowels, listeners hear a clear perceptual shift in the phonetic category (/u/ vs. /a/) around token 3, reflecting a categorical boundary along the speech continuum. (*inset*) Psychometric slope, computed from sigmoidal fits to identification functions, quantifying the “steepness” of the CP boundary, i.e., the perceptual distinction between vowel classes. (B) Speech labeling speeds (i.e., reaction times, RTs) slow near the CP boundary, suggesting that ambiguous tokens (vw 3) are more difficult to classify than within-category speech sounds (e.g., vws 1–2 or vws 4–5). Figure adapted from Bidelman et al. (2013), with permission from Elsevier.



**Fig. 2.** Neural oscillatory responses to categorical speech sounds. (A) Time–frequency maps for prototypical [vw1 = /u/, vw5 = /a/] and the ambiguous vowel tokens (vw3), distal and proximal to the CP boundary, respectively. (*top rows*) ERSPs quantify both “evoked” and “induced” changes in EEG power. (*bottom rows*) ITPCs quantify phase-locking index (PLI) across stimulus presentations and show evoked neural activity temporally consistent across trials (cf. power spectrum of the ERP). (B and C) Statistical difference maps ( $p < 0.001$ ) between vw1 and vw3 conditions. Vowels at the continuum’s endpoints (vw1) offer a clear phonetic identity, and induce increased  $\beta$ -power (13–30 Hz) in the EEG; ambiguous tokens near the CP boundary (vw3) induce an increase in higher  $\gamma$ -band (60–80 Hz) oscillations. Dotted lines demarcate the onset of time-locking speech tokens. Time–frequency regions marked ‘1’, ‘2’, and ‘3’ are described in the text.

Of particular prominence are the differences that emerge in  $\beta$  (~25 Hz, 200–300 ms) and high  $\gamma$  (50–70 Hz, 200–300 ms) band modulations between vowel conditions. Difference maps computed via permutation resampling highlight the distinction between neural responses induced by prototypical (vw1) and ambiguous (vw3) vowel stimuli (i.e.,  $vw1 > vw3$ ,  $p < 0.001$ ). Speech tokens that listeners perceive as having well-defined phonetic identity (vw1: /u/, see Fig. 1) induce a prominent increase in  $\beta$  synchronization compared to ambiguous speech tokens (vw3) shortly after stimulus onset (~200 ms) and then again ~500 ms post stimulus (Fig. 2B, region '1' and '3'). Interestingly, ambiguous speech sounds flanking the categorical boundary (vw3) induced an increase in high  $\gamma$ -oscillations in the 200–300 ms time window (Fig. 2B, region '2') that is not observed for unambiguous speech. Critically, differences in induced oscillatory responses were not observed for response-locked ERSPs, i.e., trials time-locked to listeners' motor response (see Supplemental Materials and Fig. S2), ruling out the possibility that the observed effects resulted from artifactual motor activity of listeners' button press. Together, our results delineate two main neural oscillatory patterns that unfold during the perceptual classification of speech sounds: (i) a modulation in  $\beta$ -activity that shows increased neural synchronization for well-defined phonetic speech sounds as compared to ambiguous tokens and (ii) an increased high-frequency  $\gamma$  signature generated during the perception of ambiguous speech sounds that do not conform to a single phonetic class.

### 2.3. Brain–behavior relations between neural oscillations and behavioral CP for speech

We evaluated the correspondence between neural and behavioral measures via correlational analysis. To provide a singular measure of induced neural oscillation, we aimed to identify a metric that would characterize the categorical nature of speech perception. To this end, we computed the change in brain activity between prototypical and ambiguous vowel sounds to index the differential response between categorical (phonetically sound) and non-categorical (phonetically ambiguous) speech. We used this neural metric as the brain analog to psychometric slopes which similarly provides an account of the degree of listeners' CP. Both  $\beta$ -band and  $\gamma$ -band responses were considered given that they showed significant modulations across the stimulus continuum (Fig. 2). For each condition and participant, we computed the peak  $\beta$ -perturbation response extracted from ERSPs in the early (150–250 ms) and late time windows (400–500 ms) that showed significant changes across the speech continuum (see regions '1' and '3', Fig. 2). Similarly, peak maximum  $\gamma$ -perturbations were extracted from ERSP spectrograms in the 200–300 ms time window (region '2', Fig. 2). We then computed the differential in these oscillatory metrics by calculating the change in each EEG band's response between well-defined phonetic speech sounds (vw1 = /u/) and those which were near the CP boundary whose phonetic category is perceptually ambiguous (vw3). Differential responses were then used in correlational analyses together with the slope of listeners' behavioral psychometric functions. We reasoned that if a given oscillatory band reflects behavioral CP, it would predict the degree of listeners' categorical percepts, as indexed by their psychometric slopes.

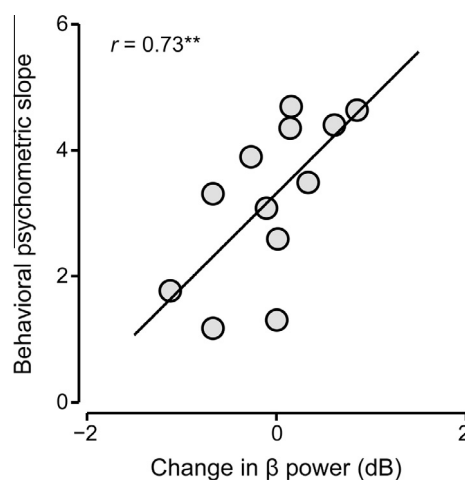
Changes in early  $\beta$ -activity did not correlate with listeners' behavioral identification performance [Spearman's rank correlation;  $r = -0.04$ ,  $p = 0.90$ ] nor did  $\gamma$ -band responses [ $r = -0.34$ ,  $p = 0.26$ ]. In stark contrast, robust correlations were observed between later  $\beta$ -activity and psychometric slopes [ $r = 0.63$ ,  $p = 0.03$ ] such that the change in  $\beta$ -power between vw1 and vw3 strongly predicted the steepness of the behavioral CP boundary. This correlation was also observed for stimulus comparisons at

the other end of the psychometric function; change in  $\beta$ -power between vw3 and vw5 similarly predicted behavioral responses [ $r = 0.52$ ,  $p = 0.042$ ]. These complementary results demonstrate that the observed correlational effect is not idiosyncratic to a specific stimulus (increasing internal/external validity) nor is it purely exogenous in nature. Rather, the robust  $\beta$ -behavior connection seems to depend on higher-order phonetic operations (i.e., endogenous processing). Given the similarity in time–frequency responses between vw1 and vw5 (Fig. 2A), we collapsed these two responses to further assess correspondence between neural  $\beta$ -modulations and behavior. We found that changes in  $\beta$ -power between phonetically salient vowels (mean vw1 and vw5) and ambiguous speech (vw3) showed even stronger correlation with listeners' psychometric slopes than when considering either endpoint in isolation [ $r = 0.73$ ,  $p = 0.009$ ] (Fig. 3). This robust brain–behavior correspondence also remained significant after adjusting for multiple correlation testing (false-discovery rate on family of tests,  $\alpha = 0.05$ ) (Benjamini & Hochberg, 1995). Collectively, these results suggest that neural  $\beta$  modulations distinguish phonetic from phonetically-ambiguous speech sounds and predict a listener's subsequent behavioral judgments of acoustic–phonetic contrasts.

Importantly, no correlations were observed between stimulus-locked  $\beta$  activity and listeners' speech labeling speeds (i.e., RTs) [early  $\beta$ :  $r = -0.07$ ,  $p = 0.83$ ; late  $\beta$ :  $r = -0.24$ ,  $p = 0.45$ ] nor between  $\gamma$  responses and RTs [ $r = 0.28$ ,  $p = 0.38$ ]. Similarly, neither early/late  $\beta$  [early  $\beta$ :  $r = 0.38$ ,  $p = 0.21$ ; late  $\beta$ :  $r = -0.06$ ,  $p = 0.83$ ] nor  $\gamma$  responses [ $r = -0.22$ ,  $p = 0.48$ ] extracted –200–100 ms prior to listeners' motor response (i.e., response-locked oscillations) showed correlations with behavior. These findings indicate that the observed oscillatory correlates of CP are related to sensory-perceptual processing and are both dissociable and distinct from motor-related processing related to the CP task.

### 3. Discussion

By measuring induced brain responses elicited during categorical speech listening task, we observed three primary findings in the current study: (1) ambiguous speech lacking a well-defined phonetic category (i.e., tokens nearest the behavioral CP boundary) induce increased modulations in the  $\gamma$ -band (>30 Hz); (2) easily



**Fig. 3.** Brain–behavior correlations underlying categorical speech perception. Changes in  $\beta$ -power between prototypical (vw1/vw5) and ambiguous (vw3) speech tokens predicts the slope of listeners' behavioral psychometric identification functions; increased neural differentiation (i.e., larger  $\beta$  change) between phonetic and phonetically-ambiguous speech sounds is associated with steeper, more categorical percepts. \*\* $p < 0.01$ .

categorized vowels (i.e., prototypical tokens at the continuum's endpoints) induce increased  $\beta$  activity (15–30 Hz) relative to phonetically-ambiguous speech sounds; (3)  $\beta$  activity predicted listeners' behavioral classification performance; larger neural activity was associated with sharper, more discrete psychometric identification functions. Neural oscillations are thought to reflect changes and information transfer and synchronization within and between neuronal networks (Giraud & Poeppel, 2012). Interpreted alongside our previous ERP study on CP (Bidelman et al., 2013), the current study extends these previous evoked response findings by revealing similar correlates of speech sound classification and linguistic operations in an alternative form of brain activity, i.e., induced neural responses.

Previous studies have demonstrated robust neural correlates of CP within various evoked brain responses, most notably the N1-P2 waves of the cortical ERPs (Bidelman et al., 2013, 2014a,b; Chang et al., 2010; Zhang et al., 2011). Converging evidence from multi-unit recordings in awake primates (Steinschneider, Fishman, & Arezzo, 2003), human fMRI (Binder et al., 2004; Zhang et al., 2011), and ERPs (Bidelman et al., 2013; Chang et al., 2010) have demonstrated that the acoustical cues necessary to establish categorical percepts are represented spatiotemporally within early superior temporal gyrus (primary/secondary auditory cortices), ~100–150 ms following the onset of sound (Bidelman et al., 2013; Chang et al., 2010; Mesgarani & Chang, 2012). These studies support the notion that early auditory cortical processing—traditionally viewed as sensory input—reflects more than external stimulus acoustics. Instead, they suggest these mechanisms reflect linguistic processing that gives rise to the abstract perceptual-phonetic attributes relevant for a listeners' intended behaviors (Bidelman et al., 2013; Chang et al., 2010; Eulitz, Diesch, Pantev, Hampson, & Elbert, 1995; Mesgarani & Chang, 2012). One interpretation is that the observed neural modulations might result from changes in the ongoing attentional state during the CP task. We feel however, that this is unlikely, given the rather uniform vigilance of our participants (speeded RTs were ~400–500 ms) and the fact that higher frequency oscillatory responses ( $\beta$ - rather than  $\alpha$ -band) covaried with listeners' psychometric slopes. The lack of statistical differences in  $\alpha$ -band argues against the notion that attentional allocation differed between conditions and underlies our current findings (Aissani et al., 2014).

We observed increased induced  $\gamma$ -band activity for ambiguous tokens which listeners were slower and more variable to classify (e.g., vw3). Increased  $\gamma$  power has traditionally been associated with local network synchronization (Giraud & Poeppel, 2012; Haenschel et al., 2000) and the formation of auditory objects (Tallon-Baudry & Bertrand, 1999). Enhanced  $\gamma$ -band responses, for example, are observed for familiar compared to unfamiliar sounds (Lenz, Schadow, Thaerig, Busch, & Herrmann, 2007), the successful detection of target vs. non-target stimuli (Gurtubay, Alegre, Labarga, Malanda, & Artieda, 2004), and when acoustic cues are matched to representations in auditory-lexical memory (Shahin et al., 2009). Under these descriptions, we would have expected prototypical vowels (at the continuum's ends) to elicit the increased  $\gamma$  responses given that they reflect more exemplary speech sounds which would converge with listeners' phonetic representations in long-term auditory memory (e.g., Shahin et al., 2009). In contrast, we found the opposite pattern: increased  $\gamma$ -power for ambiguous (cf. unfamiliar) speech tokens. Thus, it is unlikely that induced  $\gamma$  oscillations in the current study reflect stimulus familiarity or auditory template matching alone.

Alternatively, enhanced  $\gamma$ -band responses we observed for ambiguous speech sounds may reflect processing of stimulus novelty (Haenschel et al., 2000). Midpoints of a CP continuum are phonetically categorized with more uncertainty and are perceived and encoded differentially depending a listeners'

trial-by-trial perception of the identical (yet ambiguous) sound (Bidelman et al., 2013). This implies that mid-point tokens eliciting bistable percepts might be heard with a certain degree of novelty. Consequently, the observed increase in  $\gamma$  oscillation for mid-continuum stimuli (Fig. 2) may reflect the fact that these tokens are less familiar to individuals, are somewhat more novel, and subsequently induce an increase in synchronization of the high-frequency EEG (e.g., Haenschel et al., 2000). However, it should be noted that  $\gamma$  activity did not correlate with listener's psychometric functions. This implies that oscillations in this high-frequency band can emerge and operate independently regardless of perceptual outcomes. Alternatively, it is possible (but unlikely) that increased  $\gamma$  activity is somewhat epiphenomenal, reflecting microsaccadic eye movements generated during stimulus uncertainty (Melloni, Schwiedrzik, Rodriguez, & Singer, 2009; Yuval-Greenberg, Tomer, Keren, Nelken, & Deouell, 2008) (see Supplemental Discussion). These findings speak to the equivocal nature and ongoing debate regarding the functional role of  $\gamma$  activity in sensory integration and perceptual binding (speech stimuli or otherwise).

More notably, we found an increase in early (~150–250 ms)  $\beta$  oscillations when participants listened to prototypical vowels at extreme ends of the speech continuum compared to stimuli near its midpoint (Fig. 2, region '1'). This suggests that  $\beta$ -band EEG power may index the degree of similarity between the incoming stimulus and internalized speech "templates." This notion converges with findings of previous studies which have implicated early  $\beta$  enhancements to template matching between stimulus and stored speech representations in auditory memory (Rojas, Teale, Sheeder, & Reite, 2000; Shahin et al., 2009). It is possible that the increased  $\beta$  elicited by the current CP speech task indexes a similar neural comparator mechanism whereby tokens along the continuum are evaluated against their closest phonetic representation internalized by the listener. Stimuli forming a match with memory representations (e.g., vw1: /u/, vw5: /a/) would tend to generate higher power  $\beta$  responses than those which straddle a phonetic category (vw 3). This may account for the pattern of  $\beta$  changes observed in the current study during speech sound classification (e.g., Fig. 2).

In addition to these early  $\beta$  oscillations, we also observed late-peaking enhancements in  $\beta$  that were also stronger for prototypical compared to ambiguous speech (Fig. 2C, region '3'). Later  $\beta$  modulations (i.e., 400–500 ms post-stimulus timeframe) have been associated with motor-related activity and the degree of uncertainty in decision tasks requiring overt response (Senkowski, Molholm, Gomez-Ramirez, & Foxe, 2006; Tzagarakis, Ince, Leuthold, & Pellizzer, 2010). While it is possible that late  $\beta$  oscillations observed here reflect motor response uncertainty prior to speech classification, this is unlikely given that this EEG band did not correlate with behavioral RTs—a requisite of motor explanations (cf. Senkowski et al., 2006; Tzagarakis et al., 2010)—and disappeared in response-locked analysis (Fig. S2). On the contrary,  $\beta$  power was actually stronger in response to prototypical vowels than ambiguous tokens, opposite the pattern expected if these oscillations were based on motor uncertainty, which typically produce a decrease (i.e., de-synchronization) in  $\beta$  power rather than an increase (i.e., synchronization) (Tzagarakis et al., 2010). Rather, we found that late  $\beta$  was positively associated with the *psychometric slopes* of listeners' identification functions. This effect appears quite robust as this brain-behavior correlation emerged when contrasting differential responses between *either* endpoint of the continuum and ambiguous speech (i.e., both vw1–vw3 and vw5–vw3) or phonetic vs. phonetically-ambiguous tokens (i.e., vw1/5 vs. vw3). Collectively, our findings suggest that induced  $\beta$  oscillations reflect the degree of listeners' categorical speech perception (i.e., identification) independent of their decision speed or motor response, *per se* (cf. Tzagarakis et al., 2010).

The slope of listeners' psychometric CP function is often interpreted to reflect the how categorical or dichotomous a speech continuum is perceived (Bidelman et al., 2014a,b; Strouse et al., 1998; Xu et al., 2006). In the present study, we found a strong correspondence between the differential magnitude of induced  $\beta$  responses across the categorical boundary and listener's psychometric slopes. Thus, our data suggest that the degree to which a listener is able to distinguish speech sounds—in a categorical sense—is captured by differential spectrotemporal encoding in the  $\beta$  frequency band.

One limitation of the current study however, is the use of vowel stimuli. CP for speech is typically much stronger for stop consonants than vowels (Altmann et al., 2014; Pisoni, 1973). While we have no reason to believe that the current findings would not hold for other speech stimuli, the observed findings likely represent a more muted effect than what may be possible using more optimal CP stimuli (e.g., /da-/ba/ continuum). It would be of interest in future studies to examine whether induced brain responses are similarly generated when categorizing stops or other speech-like stimuli (e.g., sine wave speech).

Our data are most consistent with previous studies that have linked  $\beta$ -band neural oscillations with mechanisms associated with template matching between stimuli and internalized representations stored in auditory-lexical memory (Shahin et al., 2009). In other linguistic paradigms,  $\beta$ -activity has been associated with the detection of semantic/syntactic anomalies (Kielar, Meltzer, Moreno, Alain, & Bialystok, in press; Klimesch, Schack, & Sauseng, 2005) and auditory word processing during auditory lexical-decision tasks (Brennan, Lignos, Embick, & Roberts, 2014). In an auditory semantic priming task, Brennan et al. (2014) recently showed that related words induced increased  $\beta$ -power relative to unrelated primes during lexical decision. The authors posited that neural synchronization in the  $\beta$ -frequency range was associated with a facilitation of lexical access in early linguistic processing. Inasmuch as categorical speech perception requires a similar comparison between stimulus input and a stored lexical representations, i.e., a stored template match or "perceptual magnet" (Iverson & Kuhl, 2000), the modulations in  $\beta$  encoding we observe during CP might reflect the degree of "phonetic appropriateness" of vowel sounds and how well they match internalized speech templates.

Additionally,  $\beta$ -activity has been recently linked to the development of decisional processes in a modality general manner; induced  $\beta$  activity is observed, for example, when classifying visual motion stimuli where increases in power reflect the accumulation of evidence leading to the decision (Aissani et al., 2014). During the classification of speech, we similarly observe an increase in  $\beta$  for prototypical relative to ambiguous speech sounds (Figs. 2 and 3). Greater  $\beta$  response in the former condition may reflect a higher match to the prototypes of our listeners' vowel space and thus, increased (or more consistent) evidence for listeners to make their identification decision (cf. Aissani et al., 2014). In this regard, induced  $\beta$  activity may reflect a much broader spectral fingerprint of the EEG subserving domain-general classification/decision processes not only for speech sounds, but also stimuli in other sensory modalities.

### 3.1. Conclusions

We measured induced neuroelectric brain responses elicited during categorical speech perception. Neural oscillations in EEG  $\gamma$  activity were enhanced for phonetically-ambiguous speech tokens near the categorical boundary. These tokens were also more difficult for listeners to classify and required longer time to decide on a phonetic label. However, it is possible that increased  $\gamma$  activity at the CP boundary at least partially reflects microsaccadic eye movements due to stimulus uncertainty. Based on the current

study, the role of  $\gamma$  oscillations in categorical speech perception is inconclusive and warrants further studies using high-density multichannel EEG data for source localization.

In contrast, we found that  $\beta$  activity was enhanced for prototypical tokens (continuum endpoints) relative to phonetically-ambiguous speech sounds nearing the perceptual CP boundary. More critically, individual differences in  $\beta$  activity for stimuli across the perceptually boundary predicted perceptual performance in speech identification (but not labeling speeds); larger  $\beta$  responses were associated with steeper, more dichotomous psychometric functions. These effects were observed independent of motor activity as response-locked neural activity failed to show any reliable oscillatory effects that were observable in stimulus-locked trials. We infer that induced  $\beta$ -band oscillations within the ongoing EEG code the degree to which speech sounds match internalized phonetic representations for speech (or inversely, stimulus ambiguity).

## 4. Methods

### 4.1. Subjects

We examined oscillatory activity in 12 normal-hearing young adults (age:  $23.8 \pm 4.1$  years; 9 females) (for demographic details, see Bidelman et al., 2013). All were right-handed and reported a collegiate level of education ( $17.1 \pm 2.3$  years).

### 4.2. Stimuli and task

Electrophysiological responses were elicited by a speech vowel continuum; adjacent steps differed only in first formant frequency (Bidelman et al., 2013; see also Supplemental Material). Listeners heard 200 randomly ordered exemplars of each token and labeled them with a binary response ("u" or "a") (Bidelman et al., 2013). Interstimulus intervals were jittered randomly between 400 and 600 ms (20-ms steps, rectangular distribution) to avoid stimulus refractory effects and  $\alpha$  (~8–10 Hz) entrainment that would confound time–frequency analyses.

### 4.3. EEG time–frequency analysis

The EEG recording protocol is described in our original report (Bidelman et al., 2013; see also Supplemental Methods). Trials exceeding a stringent  $\pm 50 \mu\text{V}$  were rejected as artifacts prior to time–frequency analysis. Collapsed across stimuli, an average ( $mean \pm SD$ ) of  $130 \pm 24$  trials survived artifact rejection which is sufficient for time–frequency analysis (Shahin, Trainor, Roberts, Backer, & Miller, 2010). Time–frequency decompositions of single-trial epoched data (per vowel condition) were implemented in EEGLAB using Event-related Spectral Perturbation (ERSP) and Inter-trial Phase Coherence (ITPC) analyses (Delorme & Makeig, 2004). ERSP maps represent the changes in EEG spectral power relative to the baseline pre-stimulus period (–100 to 0 ms) and contain neural activity that is both time- and phase-locked to the eliciting stimulus (i.e., evoked activity) as well as non-phase-locked responses (i.e., induced oscillatory activity) generated by the ongoing stimulus presentation (Shahin et al., 2009, 2010). ERS-Ps were constructed by computing the change in spectral power using a 120 ms long Hanning-window sinusoidal basis function transform (i.e., fixed number of cycles across frequencies) advanced in time-steps of 5 ms across single-trial epochs (e.g., Shahin et al., 2009; Trainor, Shahin, & Roberts, 2009). The analysis bandwidth spanned 1–100 Hz (1 Hz increments) to cover the  $\theta$  – (4–8 Hz),  $\alpha$  – (9–13 Hz),  $\beta$  – (14–30 Hz), and  $\gamma$  – (>30 Hz) bands of the EEG's power spectrum. ERS-Ps were visualized by plotting

spectral perturbations in dB, computed as the log ratio between the post- and pre-stimulus power. They quantify the change in EEG power at a given time and frequency, akin to an acoustic spectrogram. In our ERSP calculations, we chose to use relatively long analysis time windows to capture low-frequency energy and achieve high spectral resolution. Longer windows may partially compromise the estimate of the null distribution as the power estimates in the baseline period may have included some influence from early evoked activity after time = 0. However, it is important to note that this would only affect the *absolute* values of the baseline power spectral estimates and would not affect the *relative* differences between stimulus conditions (cf. Kiehl et al., in press). Additionally, use of a dB scale for quantifying spectral perturbations also ensured that only relative changes were considered in the analysis.

ERSP time–frequency maps contain both “evoked” and “induced” neural activity elicited by the auditory stimulus. To identify induced responses related to categorical speech processing, it was necessary to distinguish phase-locked activity, coherent from trial-to-trial (Shahin et al., 2009, 2010; Tallon-Baudry, Bertrand, Delpuech, & Pernier, 1996). To this end, phase-coherent responses were quantified by measuring the ITPC, reflecting the power of evoked neural activity (Delorme and Makeig, 2004). The ITPC represents the degree of phase-locking at each time and frequency, measured via phase-locking index (PLI), and is equivalent to the power spectrum of the ERP. PLI varies between 0 and 1, where 0 represents complete phase-independence between trials (e.g., stochastic noise) and 1 reflects perfect phase-locking. Higher PLI in the ITPC spectrogram reflects neural responses that are more temporally aligned between trials. While ERSPs reflects both “evoked” (phase-locked) and “induced” (non-phase-locked) neural responses, the ITPC, by definition, reflects only the former. Hence, changes in neural oscillatory power that occur in ERSP but not in ITPC spectrograms are indicative of induced/intrinsic brain activity (Shahin et al., 2009).

For visualization and quantification, ERSP and ITPC maps were subjected to non-parametric bootstrap resampling to assess the reliability of the time–frequency estimates. Briefly, a surrogate data distribution was constructed by selecting spectral estimates for each trial from randomly selected latency windows in the baseline period (–100 to 0 ms) and then averaging these over many resampled surrogates. Iterating over many resamples allows one to characterize the baseline amplitude distribution whose specified percentiles can be used as significance thresholds (for details, see Delorme & Makeig, 2004). For each subject, we resampled the data  $N=1000$  times corresponding to a significance level of  $p < 0.001$ . Applying this stringent criterion allowed us to display ERSP and ITPC maps with a statistical mask whereby non-significant spectral deviations from baseline (i.e., time–frequency points not exceeding a  $p < 0.001$  threshold) were set to 0 dB. Masked ERSPs and ITPCs were computed for each participant and condition. Similarly, paired-sample bootstrapping tests ( $p < 0.001$ ) were then used to statistically compare time–frequency maps across vowel conditions (Shahin et al., 2010; Trainor et al., 2009).

#### 4.4. Behavioral data analysis

Individual vowel identification scores were fit with a two-parameter sigmoid function using standard logistic regression:  $P = 1/[1 + e^{-\beta_1(x - \beta_0)}]$ , where  $P$  is the proportion of trials identified as a given vowel,  $x$ , the step number along the stimulus continuum, and  $\beta_0$  and  $\beta_1$  the location and slope of the logistic fit estimated using nonlinear least-squares regression (Bidelman et al., 2013). The open parameter  $\beta_1$  quantifies the “steepness” (i.e., rate of change) of the psychometric function and hence, the degree of phonetic distinction heard at the categorical speech boundary

(e.g., Bidelman et al., 2014b; Xu et al., 2006). Details of the reaction time measures for this behavioral data have been reported previously (Bidelman et al., 2013).

#### Acknowledgments

The author thanks Michael Weiss and Joshua Villafuerte for their assistance in collecting the original data for this study. Portions of this work were supported by a grant from the GRAMMY® Foundation awarded to G.M.B.

#### Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version, at <http://dx.doi.org/10.1016/j.bandl.2014.11.003>.

#### References

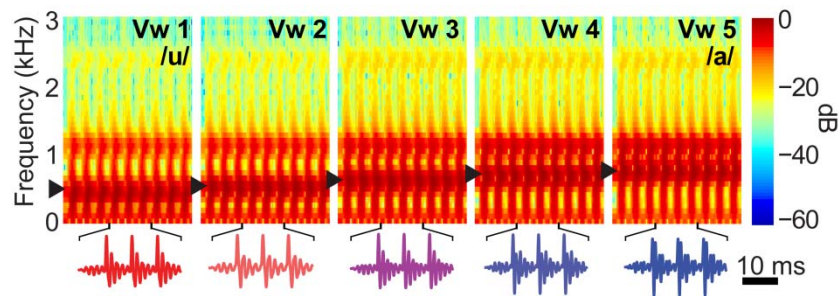
- Aissani, C., Martinerie, J., Yahia-Cherif, L., Paradis, A.-L., & Lorenceau, J. (2014). Beta, but not gamma, band oscillations index visual form–motion integration. *PLoS ONE*, 9(4), e95541.
- Altmann, C. F., Uesaki, M., Ono, K., Matsushashi, M., Mima, T., & Fukuyama, H. (2014). Categorical speech perception during active discrimination of consonants and vowels. *Neuropsychologia*, 64C, 289–300.
- 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(1), 289–300.
- Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, 79(1), 201–212.
- Bidelman, G. M., Villafuerte, J. W., Moreno, S., & Alain, C. (2014a). Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiology of Aging*, 35(11), 2526–2540.
- Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014b). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, 40, 2662–2673.
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7(3), 295–301.
- Brennan, J., Lignos, C., Embick, D., & Roberts, T. P. L. (2014). Spectro-temporal correlates of lexical access during auditory lexical decision. *Brain and Language*, 133, 39–46.
- Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., & Bressler, S. L. (2004). Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. *Proceedings of the National Academy of Sciences*, 101(26), 9849–9854.
- Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, 13(11), 1428–1432.
- Chevillet, M. A., Jiang, X., Rauschecker, J. P., & Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *Journal of Neuroscience*, 33(13), 5208–5215.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134, 9–21.
- 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 of the United States of America*, 111(19), 1–6.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303–306.
- Eulitz, C., Diesch, E., Pantev, C., Hampson, S., & Elbert, T. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *Journal of Neuroscience*, 15(4), 2748–2755.
- Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, 2(154), 1–13.
- Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–517.
- Gurtubay, I. G., Alegre, M., Labarga, A., Malanda, A., & Artieda, J. (2004). Gamma band responses to target and non-target auditory stimuli in humans. *Neuroscience Letters*, 367(1), 6–9.
- Haenschel, C., Baldeweg, T., Croft, R. J., Whittington, M., & Gruzelić, J. (2000). Gamma and beta frequency oscillations in response to novel auditory stimuli: A comparison of human electroencephalogram (EEG) data with in vitro models. *Proceedings of the National Academy of Sciences*, 97(13), 7645–7650.
- Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*, 97(1), 553–562.

- Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception and Psychophysics*, 62(4), 874–886.
- Kielar, A., Meltzer, J. A., Moreno, S., Alain, C., & Bialystok, E. (in press). Oscillatory responses to semantic and syntactic violations. *Journal of Cognitive Neuroscience*, 1–23.
- Klimesch, W., Schack, B., & Sauseng, P. (2005). The functional significance of theta and upper alpha oscillations. *Experimental Psychology*, 52, 99–108.
- 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(5044), 606–608.
- Lenz, D., Schadow, J., Thaerig, S., Busch, N. A., & Herrmann, C. S. (2007). What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. *International Journal of Psychophysiology*, 64(1), 31–38.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431–461.
- May, J. (1981). Acoustic factors that may contribute to categorical perception. *Language and Speech*, 24(3), 273–284.
- Melloni, L., Schwiedrzik, C. M., Rodriguez, E., & Singer, W. (2009). (Micro)Saccades, corollary activity and cortical oscillations. *Trends in Cognitive Sciences*, 13(6), 239–245.
- Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485, 233–236.
- Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, 20(7), 895–903.
- Pantev, C. (1995). Evoked and induced gamma-band activity of the human cortex. *Brain Topography*, 7(4), 321–330.
- Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology*, 110, 1842–1857.
- Phillips, C. (2001). Levels of representation in the electrophysiology of speech perception. *Cognitive Science*, 25(5), 711–731.
- Picton, T. W. (2010). *Human auditory evoked potentials* (1st ed.). San Diego, CA, USA: Plural Publishing.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 13(2), 253–260.
- Pisoni, D. B., & Luce, P. A. (1987). Acoustic–phonetic representations in word recognition. *Cognition*, 25(1–2), 21–52.
- Rojas, D. C., Teale, P. D., Sheeder, J. L., & Reite, M. L. (2000). Neuromagnetic alpha suppression during an auditory Sternberg task. Evidence for a serial, self terminating search of short-term memory. *Brain Research Cognitive Brain*, 101(1–2), 85–89.
- Senkowski, D., Molholm, S., Gomez-Ramirez, M., & Foxe, J. J. (2006). Oscillatory beta activity predicts response speed during a multisensory audiovisual reaction time task: A high-density electrical mapping study. *Cerebral Cortex*, 16(11), 1556–1565.
- Shahin, A. J., Picton, T. W., & Miller, L. M. (2009). Brain oscillations during semantic evaluation of speech. *Brain and Cognition*, 70(3), 259–266.
- Shahin, A. J., Trainor, L. J., Roberts, L. E., Backer, K. C., & Miller, L. M. (2010). Development of auditory phase-locked activity for music sounds. *Journal of Neurophysiology*, 103(1), 218–229.
- Steinschneider, M., Fishman, Y. I., & Arezzo, J. C. (2003). Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *Journal of the Acoustical Society of America*, 114(1), 307–321.
- Strouse, A., Ashmead, D. H., Ohde, R. N., & Grantham, D. W. (1998). Temporal processing in the aging auditory system. *Journal of the Acoustical Society of America*, 104(4), 2385–2399.
- Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, 3(4), 151–162.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *Journal of Neuroscience*, 16(13), 4240–4249.
- Trainor, L. J., Shahin, A. J., & Roberts, L. E. (2009). Understanding the benefits of musical training: Effects on oscillatory brain activity. *Annals of the New York Academy of Sciences*, 1169, 133–142.
- Tzagarakis, C., Ince, N. F., Leuthold, A. C., & Pellizzer, G. (2010). Beta-band activity during motor planning reflects response uncertainty. *Journal of Neuroscience*, 30(34), 11270–11277.
- von Stein, A., & Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: From local gamma to long range alpha/theta synchronization. *International Journal of Psychophysiology*, 38, 301–313.
- Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology*, 41(1), 48–61.
- Xu, Y., Gandour, J. T., & Francis, A. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, 120(2), 1063–1074.
- Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron*, 58, 429–441.
- Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., & Li, P. (2011). Cortical dynamics of acoustic and phonological processing in speech perception. *PLoS One*, 6(6), e20963.



### *Stimulus speech sound continuum*

A synthetic five-step vowel continuum was constructed such that each token differed minimally acoustically, yet was perceived categorically (Bidelman et al., 2013; Pisoni, 1973) (Fig. S1). This was accomplished by varying a single acoustic dimension across the stimuli, namely first formant frequency (F1). Each token was 100 ms including 10-ms of rise/fall time in order to reduce spectral splatter in the stimuli. Tokens contained identical voice fundamental (F0), second (F2), and third formant (F3) frequencies (F0: 100, F2: 1090 and F3: 2350 Hz, respectively) chosen to match prototypical productions from male speakers (Peterson & Barney, 1952). The critical stimulus variation was achieved by parameterizing first formant (F1) over five equal steps between 430 and 730 Hz such that the resultant stimulus set spanned a perceptual phonetic continuum from /u/ to /a/ (Bidelman et al., 2013). In the original experiment, stimulus presentation was controlled by a MATLAB (The MathWorks) program routed to an audiometer (GSI) via a TDT RP2 interface (Tucker-Davis Technologies) and delivered binaurally through insert earphones (ER-3A, Etymotic) at an intensity of 83 dB SPL.



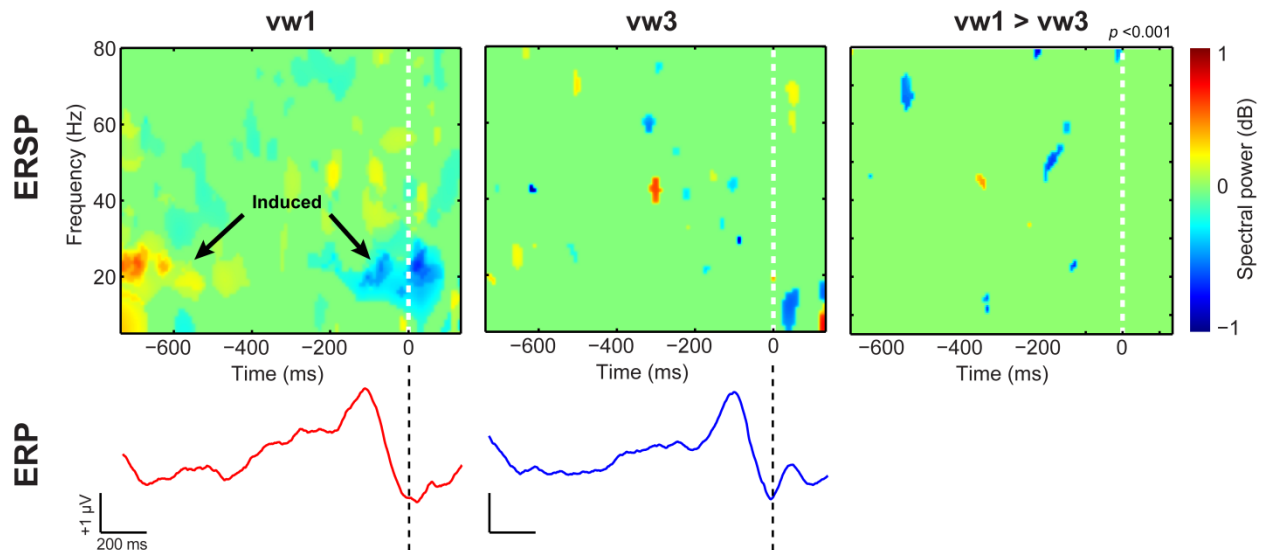
**Figure S1:** Categorical speech vowel continuum (Bidelman et al., 2013). Spectrograms of the individual 100-ms tokens; bottom insets show three periods of individual time waveforms. First formant frequency was parameterized over five equal steps from 430 to 730 Hz (arrows) such that the resultant stimulus set spanned a perceptual phonetic continuum from /u/ to /a/.

### *EEG recording protocol*

Continuous electroencephalograms were originally recorded differentially between an electrode placed on the high forehead at the hairline (~Fpz) referenced to linked mastoids; a third electrode positioned on the mid-forehead served as common ground (Bidelman et al., 2013). Inter-electrode impedance was kept below  $\leq 3$  k $\Omega$ . EEGs were digitized at 20 kHz and filtered online with a passband of 0.05–3500 Hz (NeuroScan SymAmps2). For each stimulus condition, traces were segmented (-100-600 ms epoch window) and baselined to the pre-stimulus period, and bandpass filtered between 1 and 100 Hz. Trials exceeding  $\pm 50$   $\mu$ V were rejected as artifacts (e.g., blinks) prior to averaging. Epoched data were then downsampled to 500 Hz prior to time-frequency analysis. For further details, see Bidelman et al. (2013).

### Supplemental Results

Given that participants' reaction times (RTs) for classifying speech were, on average, between 450 and 600 ms (Fig. 1B; Bidelman et al., 2013), the late emerging  $\beta$  responses may reflect activity related to listeners' motor response (e.g., Tzagarakis et al., 2010) rather than stimulus-related encoding. To rule out this possibility, we recomputed ERSPs maps time-locked to listeners' button press (i.e., RT) to directly examine induced responses generated by motor rather than stimulus activity. The analysis window encapsulated -800 ms prior to the RT and extended +200 ms after this timing-locking event. Response-locked neural activity (i.e., ERPs) revealed the classic *Bereitschaftspotential*, a readiness potential that appeared just prior to the initiation of the motor response (Kornhuber & Deecke, 1965; Sommer et al., 1994) (Fig. S2). Induced activity, as revealed by response-locked ERSPs revealed a change in  $\beta$ -band activity roughly -200–100 ms before listeners' response—coincident with the peak waves of the evoked potential. However, in stark contrast to the *increase* in late  $\beta$ -oscillatory power observed for stimulus-locked trials (cf. Fig. 2B), response-locked  $\beta$  showed an apparent de-synchronization (i.e., *decrease* in power) for prototypical (vw1) compared to ambiguous (vw3) speech. Moreover, difference maps computed via permutation resampling revealed no reliable differences in neural oscillatory activity between vowel conditions (Fig. S2, right panel). These results indicate that the oscillatory activity observed in stimulus-locked time-frequency maps (e.g., Fig. 2B) are not an artifact related to pre-motor activity, but rather, are directly related to the categorical encoding of speech sounds.



**Figure S2:** Response locked time-frequency ERS maps for a prototypical [vw1 = /u/] and the ambiguous vowel token (vw3), distal and proximal to the CP boundary (top left and middle panels). (bottom) Response locked ERP waveforms illustrate the classic *Bereitschaftspotential*, a readiness potential that appears prior to a motor response (Kornhuber & Deecke, 1965).  $\beta$ -power (13-30 Hz) decreases shortly (-200-100 ms) before the listeners' button press when classifying speech. However, induced response-locked power does not differ between the vw1 and vw3 conditions when assessed via permutation testing at the  $p < 0.001$  significance level (right panel). Dotted lines denote the time-locking response event, i.e., listeners' RT in the speech identification task.

### **Supplemental Discussion**

#### *$\gamma$ -band activity during categorical speech perception*

It is possible that larger  $\gamma$  activity may not reflect neural correlates of lexical-semantic, template matching, or novelty processing *per se*, but rather artifacts generated by microsaccadic eye movements (Melloni et al., 2009; Yuval-Greenberg et al., 2008). During stimulus uncertainty, participants often execute miniature ( $< 1^\circ$ ) involuntary eye gazes. Given that the precise timing of such events are stochastic, microsaccades can be retained in induced ERSP maps despite standard ocular artifact rejection. Trial-to-trial jitter also accounts for their ‘induced’ (rather than ‘evoked’) appearance (Yuval-Greenberg et al., 2008). While our task strictly featured the auditory modality, it is possible that when listening to ambiguous speech tokens (e.g., vw3), higher levels of stimulus uncertainty/confusion generated similar microsaccadic movements. This may explain the apparent increase in  $\gamma$  activity we observed for mid-continuum tokens (Fig. 2, region ‘2’). While  $\gamma$  oscillations in the current study do broadly agree with the temporal dynamics reported for microsaccades (i.e., both occur  $\sim 250$  ms post stimulus onset), they do not generally conform with their spectral profile;  $\gamma$  responses observed in this study carry a much more restricted bandwidth (60-80 Hz) whereas microsaccades are more broadband (30-90 Hz) in nature (Melloni et al., 2009; Yuval-Greenberg et al., 2008). Hence, our data provide no definitive evidence as to the presence/absence of microsaccadic activity. Future studies with high-density electrode montages and/or intracranial recordings are needed to fully reconcile various studies and clarify the relation between induced  $\gamma$ -band activity and cognitive computations, speech tasks or otherwise.

#### **Supplemental references:**

- Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, 79(1), 201-212.
- Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Arch*, 284, 1-17.
- Melloni, L., Schwiedrzik, C. M., Rodriguez, E., & Singer, W. (2009). (Micro)Saccades, corollary activity and cortical oscillations. *Trends in Cognitive Sciences*, 13(6), 239-245.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, 24(2), 175-184.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 13(2), 253-260.

Sommer, W., Leuthold, H., & Ulrich, R. (1994). The lateralized readiness potential preceding brief isometric force pulses of different peak force and rate of force production. *Psychophysiology*, 31(5), 503-512.

Tzagarakis, C., Ince, N. F., Leuthold, A. C., & Pellizzer, G. (2010). Beta-band activity during motor planning reflects response uncertainty. *Journal of Neuroscience*, 30(34), 11270–11277.

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient induced gamma-band response in eeg as a manifestation of miniature saccades. *Neuron*, 58, 429-441.