

AMPLIFIED INDUCED NEURAL OSCILLATORY ACTIVITY PREDICTS MUSICIANS' BENEFITS IN CATEGORICAL SPEECH PERCEPTION

GAVIN M. BIDELMAN*

Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA

School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA

University of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, TN, USA

Abstract—Event-related brain potentials (ERPs) reveal musical experience refines neural encoding and confers stronger categorical perception (CP) and neural organization for speech sounds. In addition to evoked brain activity, the human EEG can be decomposed into induced (non-phase-locked) responses whose various frequency bands reflect different mechanisms of perceptual-cognitive processing. Here, we aimed to clarify which spectral properties of these neural oscillations are most prone to music-related neuroplasticity and which are linked to behavioral benefits in the categorization of speech. We recorded electrical brain activity while musicians and nonmusicians rapidly identified speech tokens from a sound continuum. Time-frequency analysis parsed evoked and induced EEG into alpha (~10 Hz), beta (~20 Hz), and gamma (> 30 Hz) frequency bands. We found that musicians' enhanced behavioral CP was accompanied by improved evoked speech responses across the frequency spectrum, complementing previously observed enhancements in evoked potential studies (i.e., ERPs). Brain-behavior correlations implied differences in the underlying neural mechanisms supporting speech CP in each group: modulations in induced gamma power predicted the slope of musicians' speech identification functions whereas early evoked alpha activity predicted behavior in nonmusicians. Collectively, findings indicate that musical training tunes speech processing via two complementary mechanisms: (i) strengthening the formation of auditory object representations for speech signals (gamma-band) and (ii) improving network control and/or the matching of sounds to internalized memory templates (alpha/beta-band). Both neurobiological enhancements may be deployed behaviorally and account for musicians' benefits in the perceptual categorization of speech. © 2017 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: alpha activity, categorical speech perception, experience-dependent plasticity, induced oscillations, gamma activity.

INTRODUCTION

To successfully perceive auditory objects, the human brain must assemble diverse sensory information into common, well-formed groupings, a process known as categorical perception (CP). At its core, CP is known as the “invariance” or “many-to-one mapping” problem whereby an infinite collection of sensory features must be converted into a finite, invariant perceptual space to be acted upon by the perceptual system. CP is particularly evident in speech perception. When presented with a gradually morphed continuum of equidistant acoustic speech sounds, listeners' perception typically shifts abruptly near the midpoint, marking a change in the perceived category. CP is critical to speech-language abilities (Mody et al., 1997), and even though categorical boundaries emerge and are codified in early life (Eimas et al., 1971; Kuhl et al., 1992), whether or not they can be modified with training or experience has remained relatively unexplored (for cross-language and cross-domain differences in CP, see Näätänen et al., 1997; Xu et al., 2006; Bidelman and Lee, 2015; Bidelman and Walker, in press).

In this regard, musicians represent an ideal model to investigate how auditory experience alters speech listening skills and its underlying neural substrates. Recent studies have suggested that musicians have enhanced neural processing in several sensory modalities and benefits in perceptual-cognitive skills including speech and language processing (Alain et al., 2014; Moreno and Bidelman, 2014). Supporting behavioral enhancements, neurophysiological studies have revealed functional differences in brainstem and cortical neuroelectric activity in musicians in the form of larger responsiveness, higher fidelity of neural representations, and more efficient neural encoding of speech signals (Musacchia et al., 2008; Bidelman et al., 2014; Bidelman and Alain, 2015).

Extending prior studies on the benefits of musicianship to speech-language function, we recently demonstrated that musicians' speech listening skills extend to categorical processing, a higher order linguistic operation requiring a comparison between acoustic speech signals and their internalized memory

*Corresponding author. Address: School of Communication Sciences & Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA. Fax: +1-(901)-525-1282.

E-mail address: g.bidelman@memphis.edu

Abbreviations: CP, categorical perception; ERPs, Event-related brain potentials; ITPC, inter-trial phase coherence; NMs, nonmusicians; PLI, phase-locking index.

representations (i.e., match to “phonetic template”) (Bidelman et al., 2014). Behaviorally, musicians demonstrated faster classification and more pronounced (i.e., steeper) psychometric identification functions for speech, indicating enhancements in mapping sound objects to their categorical identities. Speech-evoked ERPs further revealed that musicians’ behavioral benefits were supported by a concert of neuroplastic effects in speech encoding from brainstem to cortex (Bidelman et al., 2014; Bidelman and Alain, 2015). While these and other studies reveal early *evoked* activity underlying speech coding and how it is shaped by experience, they cannot speak to putative connections between *induced* brain responses that may also underlie skilled categorization (cf. Bidelman, 2015) nor how (if) such intrinsic brain activity is altered in an experience-dependent manner (e.g., Trainor et al., 2009).

The EEG can be described as being either “evoked” or “induced” to task manipulations (Pfurtscheller and Lopes da Silva, 1999; Shahin et al., 2009). Evoked, phase-locked responses are obtained by cross-trial averaging to derive the conventional ERP. Additionally, ongoing neural oscillatory “rhythms” not phase-locked to stimuli can still be induced by task processing or timed stimulus events. Induced spectral measures (e.g., α , β , γ -band) are only observed via time–frequency analysis (Tallon-Baudry and Bertrand, 1999). This technique acts to circumvent the temporal jitter that normally precludes their visibility in conventional ERPs. Induced activity complements the evoked ERP by providing a window into the dynamics of speech-language function and its neural mechanisms from the perspective of intrinsic brain function. Indeed, heightened power in the γ frequency range (30–120 Hz) has been linked to synchronization of nearby brain regions (Giraud and Poeppel, 2012), auditory object construction (Tallon-Baudry and Bertrand, 1999), and semantic processing (Shahin et al., 2009). β -band (15–30 Hz) has been linked to operations related to template matching (Shahin et al., 2009) and working memory (Bashivan et al., 2014; Rose et al., 2016). More recently, we have shown that induced β and high-frequency γ rhythms accompany perceptual confusions when categorizing speech sounds, indicating a role of induced oscillations in CP (Bidelman, 2015)—at least in nonmusicians. Lastly, lower frequencies, including α -band (9–13 Hz) oscillations, have been linked to the intelligibility of speech (Becker et al., 2013) and selective inhibition of irrelevant cues (Strauß et al., 2014) when objects need to be ignored or selected against within the attentional spotlight (Foxe and Snyder, 2011). Given that various frequency components of the EEG relate to different mechanisms in speech perception, the current study aimed to further elaborate the dynamics of these brain rhythms, ascertain which spectral characteristics are associated with speech categorization, and identify which are susceptible to the neuroplastic effects of musical training.

It is well established that musicians show stronger neural encoding of musically relevant sounds as indexed by the auditory ERPs (e.g., Shahin et al., 2003; Baumann et al., 2008; Bidelman, 2013). However, recent studies have also shown that musicians have enhanced

neural encoding of speech (e.g., Musacchia et al., 2008; Bidelman and Krishnan, 2010; Parbery-Clark et al., 2011; Bidelman et al., 2014; Tierney et al., 2015). In the current study, we extend these previous ERP results by evaluating time–frequency measures of the EEG during speech categorization, a process requiring a listener to reconcile the acoustic speech signal with a phonetic template. Time–frequency differences between musicians and nonmusicians have been reported in response to non-speech sounds (e.g., pure tones and musical sounds; Shahin et al., 2008, 2010; Trainor et al., 2009). Yet, to our knowledge, the effects of musicianship on oscillatory brain activity have not been reported for speech, particularly during CP. Consequently, spectral analyses were expected to provide new insight into the physiological correlates and experience-dependent plasticity of CP by revealing group differences in *induced* brain responses and band-specific oscillations. We hypothesized that (i) musicians would show stronger auditory induced responses than musically naïve participants; and (ii) musicians’ benefits in speech CP observed in previous ERP studies would be linked to improved neural coding, particularly in higher frequency induced activity (e.g., β and/or γ oscillations). Our predictions were based on previous studies showing that musical training strengthens induced responses to non-speech auditory stimuli (Shahin et al., 2008, 2010) and that successful speech CP is dominated by coding in β and γ oscillations, EEG bands thought to carry information regarding auditory object formation and matching to internal speech templates (Shahin et al., 2009; Bidelman, 2015).

EXPERIMENTAL PROCEDURES

The current report represents a reanalysis of data from our previous ERP study (Bidelman et al., 2014) to examine experience-dependent plasticity in the *induced* components in musicians and nonmusicians.

Participants

Twenty-four young adults participated in the experiment: 12 English-speaking musicians (8 female) and 12 nonmusicians (8 female). Briefly, musicians (Ms) had received ≥ 7 years of private instruction on their principal instrument (13.6 ± 4.5 yrs) before age 13 (7.7 ± 3.5 yrs). Nonmusicians (NMs) had less than two years of musical instruction in their lifetime (0.4 ± 0.7 yrs). Each listener was right-handed (Oldfield, 1971) and had normal (< 25 dBHL) audiometric thresholds through 4000 Hz. Groups were otherwise matched in age (M: 23.8 ± 0.2 yrs, NM: 24.8 ± 2.7 yrs) and education (for complete demographic details, see Bidelman et al., 2014).

Stimuli and behavioral paradigm

We recorded speech ERPs to tokens along a 5-step vowel continuum where each step varied in first formant frequency (see Fig 1. in Bidelman et al., 2014). A total of 200 trials of each speech stimulus were presented to the listeners. They were required to label each token with

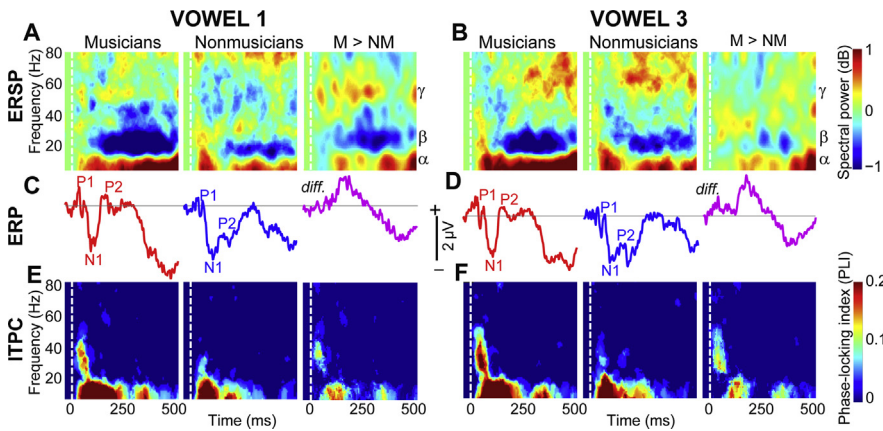


Fig. 1. Neural oscillatory responses to categorical speech sounds are stronger in musicians. Time-frequency maps for the prototypical [vw1 = /u/] and ambiguous vowel token (vw3), distal and proximal to the CP boundary, respectively. (A, B) ERSPs quantify both “evoked” and “induced” changes in EEG power relative to baseline period ($p < 0.001$). Difference maps contrasting groups are shown in rightmost columns. (C, D) Evoked potentials illustrating the obligatory cortical response reported in Bidelman et al. (2014) and the difference wave between groups (“diff.”). (E, F) ITPCs quantify phase-locking index (PLI) across stimulus presentations and show evoked neural activity temporally consistent across trials (i.e., the power spectrum of the ERP). Group differences in induced EEG energy are most prominent for prototypical speech sounds and appear in α - (~8–13 Hz), β - (13–30 Hz) and γ -band (> 30 Hz) oscillations. Dotted lines demarcate the onset of time-locking speech tokens.

a binary response (“u” or “a”). Interstimulus interval between consecutive tokens was randomly jittered uniformly between 400–600 ms to avoid unrelated α -band entrainment and listeners anticipating their response.

Time-frequency analysis

Electrophysiological procedures are fully detailed in Bidelman et al. (2014). EEGs were recorded using a vertical electrode montage (i.e., ~Fpz–A1/A2) with the mid-forehead serving as ground. Epochs in which voltages exceeded $\pm 50 \mu\text{V}$ were rejected as artifacts. Time-frequency spectrograms of each response epoch were extracted via event-related spectral perturbation (ERSP) and inter-trial phase coherence (ITPC) analyses (Delorme and Makeig, 2004) using previously published procedures (Bidelman, 2015). Briefly, ERSPs show spectral power changes over the epoch’s time course. ERSPs were calculated as the change in power (re. baseline period) at frequencies between 1 and 100 Hz (1 Hz step). A 120-ms Hanning-window was used to compute the Fourier transform and then advanced in steps of 5 ms across single-trial epochs (e.g., Shahin et al., 2009; Bidelman, 2015). This bandwidth covered α - (9–13 Hz), β - (14–30 Hz), and γ - (> 30 Hz) bands of the EEG’s spectrum. The log ratio between the post/pre-stimulus power (in dB) was then used to plot ERSPs for each time-frequency point.

ERSP spectrograms comprise both “evoked-” and “induced neural activity.” Evoked activity was isolated by analyzing activity that was coherent from trial-to-trial via ITPC maps (Tallon-Baudry et al., 1996; Shahin et al., 2009, 2010). ITPCs reflect the amount of phase-locking at each time–frequency pair, quantified by phase-locking index (PLI). PLI values vary from 0 to 1, where 0 reflects complete random trial-to-trial variation and 1, perfect trial-

to-trial dependence (i.e., temporally aligned, phase-locked activity). ERSPs reflect both “phase-locked (evoked)” and “non-phase-locked (induced)” neural responses, whereas ITPCs capture only evoked components. Thus, neural signatures appear in ERSPs and not ITPCs reveal induced neural oscillations.

We evaluated the reliability of ERSP and ITPC spectrograms via bootstrapping. A surrogate data distribution was created by randomly selecting portions of the baseline (–100 to 0 ms) and then averaging across resamples. This procedure resulted in an estimate of the baseline’s amplitude distribution whose percentiles can be used as significance thresholds (Delorme and Makeig, 2004). We used $N = 1000$ resamples to achieve a statistical threshold of $p < 0.001$. Adopting this criterion allowed us to mask ERSPs/ITPCs so that power changes within the post-stimulus epoch that were not significantly different from the baseline were set to 0 dB. Masked ERSPs and ITPCs were computed for each subject and stimulus.

Behavioral data analysis

We modeled each participants’ identification scores using a standard two-parameter sigmoid: $P = 1/[1 + e^{-\beta_1(x - \beta_0)}]$, where P is the proportion of identified trials, x the step number along the continuum, and β_0 and β_1 are the function’s location and slope estimated using nonlinear least-squares fitting (Bidelman et al., 2014). The parameter β_1 quantifies the “steepness” of the psychometric function and thus, how strong a listener differentiates categories at the perceptual boundary (Xu et al., 2006; Bidelman et al., 2014). Comparing β_1 parameters between musicians and nonmusicians was used to reveal differences in the sharpness of the CP boundary between groups (Bidelman et al., 2014).

RESULTS

ERSPs and ITPCs for a prototypical (vw1) and ambiguous (vw3) token are illustrated in Fig. 1, and provided an initial qualitative view of group differences across the EEG’s spectrum. For reference, the evoked potentials (ERPs) are shown in panels C–D of the figure (ERP results are reported in Bidelman et al., 2014). Neural oscillatory signatures present in ERSPs but not ITPCs reveal “induced” activity not phase- but nonetheless time-locked to the speech stimulus and/or identification task. Difference spectrograms contrast brain activity between musicians and nonmusicians (i.e., M > NM).

Speech stimuli perceived with a defined identity (vw1: /u/) induced decreased power in the β -band in musicians relative to nonmusicians ~350–400 ms post stimulus.

This group difference appeared weaker for ambiguous speech tokens (vw3) and was largely circumscribed to later portions of this time window (i.e., ~ 375 ms). Musicians' responses also showed short segments of increased synchronization in α - (~ 125 ms) and γ -band (~ 300 ms) oscillations compared to their nonmusician peers. Together, our initial view of the data revealed three main differences in the pattern of induced neural activity in musicians and nonmusicians during auditory vowel categorization including modulations in α -, β -, and γ -spectral activity.

To quantify these group effects, we extracted the time course of spectral responses in the α -, β -, and γ -bands. Fig. 2 shows group averaged time courses of ITPC (A) and ERSP (B) responses elicited by the prototypical vw1 (/u/) token extracted from the spectrographic maps shown in Fig. 1A. ITPC waveforms index phase-locked components of neural activity that remain phase-consistent across trials, whereas ERSP waveforms also index the time-varying changes in intrinsic (non-phase-locked) spectral power of the EEG. A running t -test (Guthrie and Buchwald, 1991) was used to assess sample-by-sample differences between each group's neural signature ($p < 0.05$, $N = 5000$ bootstrapped resamples). We required segments persist contiguously for ≥ 15 ms to be considered reliable and help control false positives (Guthrie and Buchwald, 1991; Chung and Bidelman, 2016).

In response to the prototypical vowel (vw1), band-specific time courses revealed musicians' responses contained more robust evoked activity in the α -, β -, and

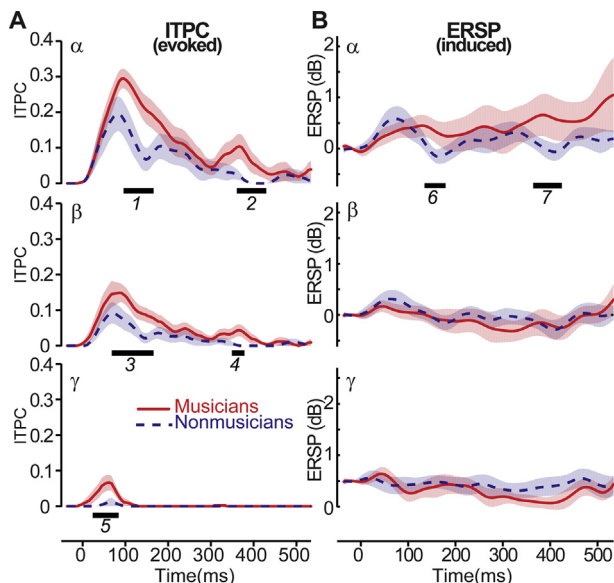


Fig. 2. Evoked and induced brain activity is enhanced in musicians. (A) *Evoked* response time courses for each group extracted from ITPC maps (see Fig. 1) index the phase-locked components of neural activity. (B) *Induced* time courses for each group extracted from ERSP maps index the time-varying changes in spectral power of intrinsic (non-phase-locked) neural activity. A running t -test (Guthrie and Buchwald, 1991) shows contiguous segments of at least 15 ms duration where spectral power differs between groups (■ segments; $p < 0.05$, $N = 5000$ bootstrap resamples). All waveforms reflect responses to the prototypical /u/ vowel (vw1). For reference to the numbered segments, see text. Shaded regions = ± 1 s.e.m.

γ -bands (see time-intervals “1–5”, Fig. 2A). Specifically, musicians' ITPC α - and β -band responses were larger in the ~ 100 -ms and ~ 400 -ms time windows and ~ 50 – 100 -ms window for the γ -band. Fewer group differences were observed in ERSP activity. Yet, stronger induced α activity was observed in musicians' responses roughly 100 ms and 400 ms after speech onset (see time-intervals “6–7”; Fig. 2B). A similar pattern of group effects was observed when considering time courses for tokens at the midpoint (vw3) and opposite end (vw5) of the speech continuum with additional group differences (i.e., $M > NM$) emerging in the early (150 ms) and later (500 ms) β -band of the induced ERSP for vw3 and γ -band (~ 200 ms) for vw5 (data not shown). Together, these findings demonstrate that differences in neural activity to speech between groups is driven by specific temporal patterns of change in underlying α , β , and γ evoked and induced activity.

Brain-behavior relations

Fig. 3A illustrates psychometric identification functions for musicians and nonmusicians. Of interest is the slope of these functions (Fig. 3B) which represents the degree to which listeners dichotomize the two vowel classes of our speech continuum (Bidelman et al., 2014; Bidelman and Lee, 2015). Musicians showed much steeper CP boundaries ($t_{22} = 2.45$, $p = 0.023$), indicating stronger, more dichotomous CP than their nonmusician peers (Fig. 3B).

Correlations between induced brain responses and psychometric slopes assessed putative relations between ongoing neural oscillations of the EEG and behavioral CP (Fig. 3C). From each time segment and band (α , β , and γ) showing significant group effects in any of the vowel tokens (e.g., time-intervals “1–7”, Fig. 2), we derived a singular measure of categorical neural coding by evaluating the change in power between prototypical and ambiguous stimuli (i.e., vw1-vw3) (Bidelman, 2015). For these analyses, responses were extracted in the temporal center of each segment for the vw1 and vw3 response and their difference computed. This differential metric has been used previously to predict the strength of individuals' CP, as indexed by the slope of the psychometric function (e.g., Bidelman, 2015).

In addition to bands showing group differences, we also extracted ITPC/ERSP responses in early γ -band (75–100 ms) in light of our specific hypotheses and putative link between this frequency range and speech categorization (Bidelman, 2015) and the formation of perceptual objects (Tallon-Baudry and Bertrand, 1999; Trainor et al., 2009). This early window was selected because this was a common time interval that showed the majority of group differences across ERSPs/ITPCs waveforms (e.g., “1”, “3”, “5”, “6”, Fig. 2) and vowel tokens.

We found a significant correlation between changes in early induced γ -band ERSP and the psychometric slopes for musicians (Pearson's $r = 0.62$, $p = 0.032$) (Fig. 3C). This same correspondence was not observed in the nonmusician group ($r = -0.43$, $p = 0.16$). The relation between brain and behavior measures was also stronger in musically trained ears (Fisher transform:

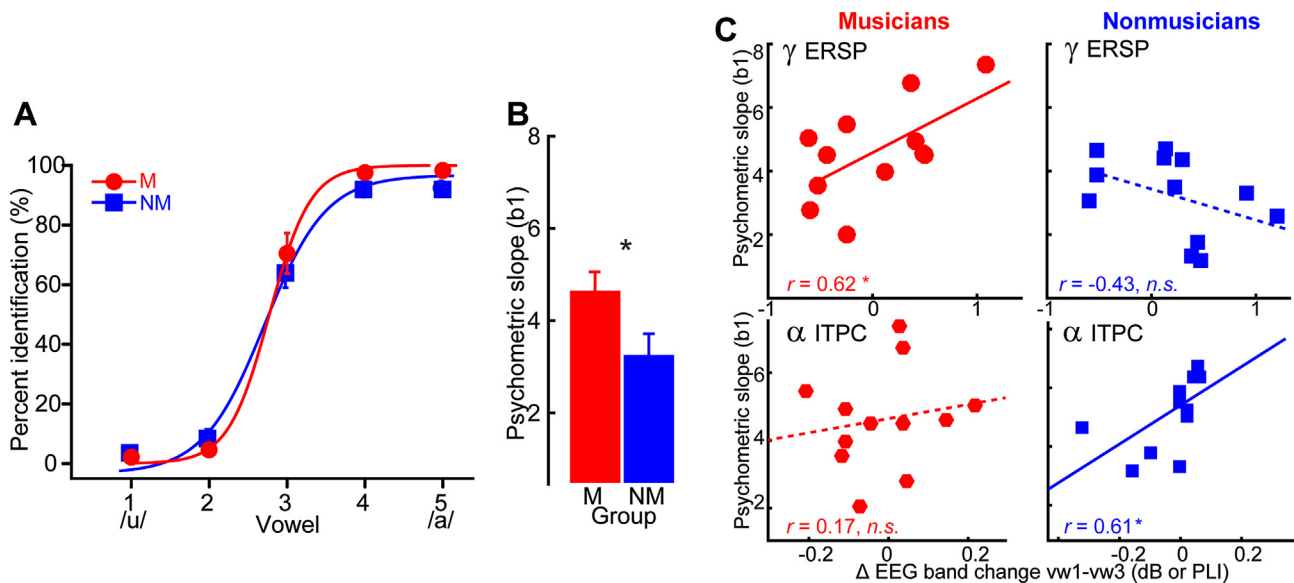


Fig. 3. Brain-behavior correspondence between induced neural oscillations and categorical speech perception is enhanced in musicians. (A) Psychometric functions for identifying vowel stimuli and their corresponding (B) slopes, estimated from the β_1 slope parameter of sigmoidal fits. Behavioral identification functions reveal sharper, more dichotomous speech classification in musically trained listeners. (C) In musicians, increased neural differentiation between prototypical (vw1) and ambiguous (vw3) speech tokens in the ERSP γ -band is associated with steeper, more categorical percepts. In contrast to musicians, neural differentiation in the ITPC α -band activity predicts nonmusicians' behavioral CP. $p < 0.05$; errorbars = ± 1 s.e.m.; solid lines = significant correlations, dotted lines = insignificant correlations.

$z = 2.50$, $p = 0.0123$). A complementary correlation occurred when we compared stimuli at the opposite continuum endpoint. That is, changes in γ -band amplitude from vw5 to vw3 also predicted the sharpness of behavioral identification for musicians ($r = 0.58$, $p = 0.0465$). This correlation was not observed in nonmusicians ($r = 0.21$, $p = 0.10$). The complementary nature of the results demonstrates that musicians' correlational link is driven by phonetic properties of speech perception (i.e., abstract categories) rather than mere stimulus acoustics which are different between ends of continuum (Bidelman, 2015). In contrast to musicians, nonmusicians' behavioral CP was instead predicted ($r = 0.61$, $p = 0.037$) by neural differentiation in early evoked ITPC α -band (e.g., time-interval "1", Fig. 2A) which was not observed in musicians ($r = 0.17$, $p = 0.59$) (Fig. 3C). However, the difference between these group correlations was not significant ($z = -1.13$, $p = 0.26$). No other correlations reached significance (all $ps > 0.05$).

In summary, we found that the strength of musicians' categorical speech percepts was predicted by early neural differentiation in induced γ -band power whereas nonmusicians' was predicted by early evoked α in the same time period. Still, some caution is warranted when interpreting the strength of these correlations as they became marginal ($p < 0.09$) or did not survive adjustments when we corrected for multiple comparisons (Holm, 1979). However, this might be expected given the large number of correlational tests and smaller sample size of our cohort. Notwithstanding the strong group effects in the neural data alone, these correlational results suggest a putative link between spectral EEG activity and behavioral CP as well as mechanistic differences between groups to explore further in future studies.

DISCUSSION

The present study measured induced neural activity to assess auditory categorical speech processing in musicians and nonmusicians. Results show: (i) enhanced intrinsic oscillatory responses to speech sounds in musically trained listeners, (ii) heightened early/late evoked α - β band (α : 150/400 ms; β : 100/250 ms) and γ -band (100 ms) power, and (iii) evidence of potential differences in the brain mechanisms supporting speech categorization between groups with early induced γ -activity driving behavior in musicians but early evoked α -activity driving CP in nonmusicians. Interpreted alongside our prior ERP studies on musical training and CP (Bidelman et al., 2014), current findings corroborate and extend previous evoked response findings by revealing similar neuroplastic benefits of musical training on speech processing within both evoked and induced brain activity and improved behavior.

Various spectral bands of the induced EEG spectrum are thought to reflect different mechanisms of perceptual-cognitive function (Giraud and Poeppel, 2012). While still debated, current conceptualizations of induced brain activity assume that lower bands of oscillation reflect long-range network communication and cognitive processing (e.g., working memory, attentional deployment, etc.) whereas higher frequency activity reflects local synchronization in sensory cortices that tags stimulus properties and forms the basis of perceptual objects (von Stein and Sarnthein, 2000). In relation to communication, neural oscillations are thought to characterize how the brain temporally organizes the incoming speech signal (Giraud and Poeppel, 2012). Under this framework, α oscillations have been associated with attentional filtering and speech intelligibility (Shahin et al., 2009; Foxe and

Snyder, 2011; Becker et al., 2013; Strauß et al., 2014), β oscillations with speech template matching (Shahin et al., 2009; Bidelman, 2015), and γ oscillations with the formation of perceptual objects and the integration of top-down and bottom-up processes (Tallon-Baudry and Bertrand, 1999; Trainor et al., 2009).

It is notable then that we found group differences in evoked spectral responses in all three of these bands which represent unique speech mechanisms. However, we also found they followed a time dependence that was inversely related to frequency (Fig. 2). That is, early group differences were observed in higher frequency evoked γ oscillations \sim 50–100 ms, followed by enhancements in mid-frequency β (75–150 ms), ending with lower frequency α increases in early (100–150 ms) and later (\sim 400 ms) time windows. In this regard, the time course of these new spectral results extend previous observed enhancements in musicians' auditory cortical ERPs to speech (Baumann et al., 2008; Musacchia et al., 2008; Bidelman and Alain, 2015; Tierney et al., 2015) by parsing those evoked effects in a frequency-specific manner. Musicians' β -band enhancements are particularly notable in light of the fact that this frequency range is thought to reflect speech template matching and the formation of phonetic categories (Shahin et al., 2009; Bidelman, 2015). Stronger internalized speech templates could offer musicians a more accurate representations of the speech signal and enable better categorization observed in this group (e.g., Fig. 3A). Additionally, we found that musicians' evoked spectral enhancements were accompanied by early increases in *induced* α activity \sim 100–150 ms after speech onset. We interpret this latter result as reflecting possible improvements in musicians' early automatic allocation of attention to the incoming speech signal in preparation for comparison to internalized auditory-speech templates (e.g., Shahin et al., 2009). Previous ERP studies have similarly suggested enhanced attentional processing and automatic reorienting in musicians (Chobert et al., 2011; Zendel and Alain, 2014; Bidelman and Alain, 2015). Thus, in relation to the various mechanisms and time course of speech processing, our results provide evidence that the representation of speech sound objects (γ -band), reconciliation with speech templates (β -band), and early attentional allocation (α -band) are all enhanced in musically trained individuals.

Additional differences were observed in how neural rhythms predicted each groups' behavioral CP. In musicians, modulations in γ -band were associated with improved speech identification; γ -power was stronger for phonetically salient (vw1) compared to ambiguous speech (vw3) of the continuum suggesting that this band distinguished prototypical and phonetically ambiguous speech sound objects. Interestingly, a larger response differential between tokens at either of the two continuum endpoints (vw1 or vw5) relative to the midpoint (vw3) was associated with musicians' stronger behavioral CP (i.e., steeper psychometric functions) (Fig. 3C). However, this gamma-behavior relation was only present in musicians. These results corroborate our previous report that similarly failed to find an association

between induced γ activity and listeners' psychometric functions in nonmusician listeners (Bidelman, 2015). We have previously posited that induced γ may reflect perceptual confusions for ambiguous sounds not conforming to a singular phonetic identity—as is the case for tokens (vw3) near the perceptual boundary (Bidelman, 2015). Consequently, larger γ differentiation only in musicians may reflect their more authentic depiction of speech sound categories relative to those which do not carry an obvious identity.

Contrastively, CP in nonmusicians was instead predicted by early evoked α -activity (Fig. 3C). Although preliminary in nature, our correlational data suggest that speech CP might potentially be driven by different neural mechanisms (cf. task strategy) in highly trained vs. untrained listeners. Tentatively, we posit that successful CP might be governed by more automatic, or lower level processes in musicians (e.g., more accurate auditory template matching indexed by γ modulations) but higher cognitive control mechanisms in nonmusicians (as indexed by α modulations) (for similar effects in older musicians, see Bidelman and Alain, 2015). This is further suggested by the fact that nonmusicians' α -activity was weaker than musicians' across the board (Fig. 2), possibly indicating a less successful deployment of (automatic) attentional resources when identifying speech. Alternatively, weaker α -activity has been linked to poorer maintenance and capacity of working memory during speech processing (Bashivan et al., 2014), suggesting another interpretation of the data.

Collectively, findings indicate that while multiple bands of the EEG differentiate musicians' and nonmusicians' neural encoding of speech, it is primarily higher frequency oscillations (i.e., γ -band) that drive musicians' experience-dependent enhancements in CP at the behavioral level. These findings support the general notion that *evoked* γ signatures index physical sound properties which can be modulated by listening experience, whereas *induced* γ activity reflects higher perceptual learning that links neural representation to a perceptual output (Shahin et al., 2008). Under this notion, musicians' enhanced neural responsivity for speech, as indexed by ERP studies (e.g., Musacchia et al., 2008; Parbery-Clark et al., 2011; Bidelman et al., 2014; Tierney et al., 2015), while presumably necessary, may not produce perceptual benefits in speech-listening skills in and of itself (Tervaniemi et al., 2009; Bidelman et al., 2011). Rather, sensory representations for speech must be bound into a perceptual object that is then selected for perceptual enhancement and acted upon behaviorally. Given that induced γ -band oscillations are implicated in constructing auditory object representations and perceptual binding (Pfurtscheller and Lopes da Silva, 1999; Tallon-Baudry and Bertrand, 1999; von Stein and Sarnthein, 2000), it is conceivable that *induced* γ signatures of the brain's speech networks must also be tuned alongside evoked properties of neural activation to produce musicians' enhanced speech-listening skills observed in this and previous studies (Musacchia et al., 2008; Bidelman et al., 2014; Bidelman and Alain, 2015; Tierney et al., 2015).

Acknowledgments—Kelsey Mankel kindly provided comments on previous versions of this article. Portions of this work were supported by a grant from the GRAMMY Foundation® awarded to G.M.B.

REFERENCES

- Alain C, Zendel BR, Hutka S, Bidelman GM (2014) Turning down the noise: The benefit of musical training on the aging auditory brain. *Hear Res* 308:162–173.
- Bashivan P, Bidelman GM, Yeasin M (2014) Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity. *Eur J Neurosci* 40:3774–3784.
- Baumann S, Meyer M, Jancke L (2008) Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J Cogn Neurosci* 20:2238–2249.
- Becker R, Pefkou M, Michel C, Hervais-Adelman A (2013) Left temporal alpha-band activity reflects single word intelligibility. *Front Sys Neurosci* 7.
- Bidelman GM (2013) The role of the auditory brainstem in processing musically-relevant pitch. *Front Psychol* 4:1–13.
- Bidelman GM (2015) Induced neural beta oscillations predict categorical speech perception abilities. *Brain Lang* 141:62–69.
- Bidelman GM, Alain C (2015) Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J Neurosci* 35:1240–1249.
- Bidelman GM, Gandour JT, Krishnan A (2011) Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain Cogn* 77:1–10.
- Bidelman GM, Krishnan A (2010) Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res* 1355:112–125.
- Bidelman GM, Lee C-C (2015) Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *Neuroimage* 120:191–200.
- Bidelman GM, Walker B (in press) Attentional modulation and domain specificity underlying the neural organization of auditory categorical perception. *Eur. J. Neurosci.* <http://dx.doi.org/10.1111/ejn.13526>.
- Bidelman GM, Weiss MW, Moreno S, Alain C (2014) Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur J Neurosci* 40:2662–2673.
- Chobert J, Marie C, Francois C, Schon D, Besson M (2011) Enhanced passive and active processing of syllables in musician children. *J Cogn Neurosci* 23:3874–3887.
- Chung W-L, Bidelman GM (2016) Cortical encoding and neurophysiological tracking of intensity and pitch cues signaling English stress patterns in native and nonnative speakers. *Brain Lang* 155–156:49–57.
- Delorme A, Makeig S (2004) EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *J Neurosci Meth* 134:9–21.
- Eimas PD, Siqueland ER, Jusczyk P, Vigorito J (1971) Speech perception in infants. *Science* 171:303–306.
- Foxe JJ, Snyder AC (2011) The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front Psychol* 2:1–13.
- Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat Neurosci* 15:511–517.
- Guthrie D, Buchwald JS (1991) Significance testing of difference potentials. *Psychophysiology* 28:240–244.
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70.
- Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255:606–608.
- Mody M, Studdert-Kennedy M, Brady S (1997) Speech perception deficits in poor readers: Auditory processing or phonological coding? *J Exp Child Psychol* 64:199–231.
- Moreno S, Bidelman GM (2014) Understanding neural plasticity and cognitive benefit through the unique lens of musical training. *Hear Res* 308:84–97.
- Musacchia G, Strait D, Kraus N (2008) Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear Res* 241:34–42.
- Nääätänen R, Lehtokoski A, Lennes M, Cheour M, Huottilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385:432–434.
- Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Parbery-Clark A, Marmel F, Bair J, Kraus N (2011) What subcortical-cortical relationships tell us about processing speech in noise. *Eur J Neurosci* 33:549–557.
- Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clin Neurophysiol* 110:1842–1857.
- Rose NS, LaRocque JJ, Riggall AC, Gosseries O, Starrett MJ, Meyerling EE, Postle BR (2016) Reactivation of latent working memories with transcranial magnetic stimulation. *Science* 354:1136–1139.
- Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE (2003) Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J Neurosci* 23:5545–5552.
- Shahin AJ, Picton TW, Miller LM (2009) Brain oscillations during semantic evaluation of speech. *Brain Cogn* 70:259–266.
- Shahin AJ, Roberts LE, Chau W, Trainor LJ, Miller LM (2008) Music training leads to the development of timbre-specific gamma band activity. *Neuroimage* 41:113–122.
- Shahin AJ, Trainor LJ, Roberts LE, Backer KC, Miller LM (2010) Development of auditory phase-locked activity for music sounds. *J Neurophysiol* 103:218–229.
- Strauß A, Wöstmann M, Obleser J (2014) Cortical alpha oscillations as a tool for auditory selective inhibition. *Front Hum Neurosci* 8:350.
- Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci* 3:151–162.
- 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. *J Neurosci* 16:4240–4249.
- Tervaniemi M, Kruck S, De Baene W, Schroger E, Alter K, Friederici AD (2009) Top-down modulation of auditory processing: Effects of sound context, musical expertise and attentional focus. *Eur J Neurosci* 30:1636–1642.
- Tierney AT, Krizman J, Kraus N (2015) Music training alters the course of adolescent auditory development. *Proc Natl Acad Sci U S A* 112:10062–10067.
- Trainor LJ, Shahin AJ, Roberts LE (2009) Understanding the benefits of musical training: Effects on oscillatory brain activity. *Ann N Y Acad Sci* 1169:133–142.
- von Stein A, Sarnthein J (2000) Different frequencies for different scales of cortical integration: From local gamma to long range alpha/theta synchronization. *Int J Psychophysiol* 38:301–313.
- Xu Y, Gandour JT, Francis A (2006) Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J Acoust Soc Am* 120:1063–1074.
- Zendel BR, Alain C (2014) Enhanced attention-dependent activity in the auditory cortex of older musicians. *Neurobiol Aging* 35:55–63.