

COGNITIVE NEUROSCIENCE

Attentional modulation and domain-specificity underlying the neural organization of auditory categorical perception

Gavin M. Bidelman^{1,2,3}  and Brea S. Walker^{1,4}¹Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA²School of Communication Sciences & Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA³Department of Anatomy and Neurobiology, University of Tennessee Health Sciences Center, Memphis, TN, USA⁴Department of Psychology, University of Memphis, Memphis, TN, USA

Keywords: auditory attention, auditory event-related potentials, categorical perception, domain-specificity, musical pitch perception, music-language relations

Edited by John Foxe

Received 17 October 2016, revised 13 January 2017, accepted 13 January 2017

Abstract

Categorical perception (CP) is highly evident in audition when listeners' perception of speech sounds abruptly shifts identity despite equidistant changes in stimulus acoustics. While CP is an inherent property of speech perception, how (if) it is expressed in other auditory modalities (e.g., music) is less clear. Moreover, prior neuroimaging studies have been equivocal on whether attentional engagement is necessary for the brain to categorically organize sound. To address these questions, we recorded neuroelectric brain responses [event-related potentials (ERPs)] from listeners as they rapidly categorized sounds along a speech and music continuum (active task) or during passive listening. Behaviorally, listeners' achieved sharper psychometric functions and faster identification for speech than musical stimuli, which was perceived in a continuous mode. Behavioral results coincided with stronger ERP differentiation between prototypical and ambiguous tokens (i.e., categorical processing) for speech but not for music. Neural correlates of CP were only observed when listeners actively attended to the auditory signal. These findings were corroborated by brain-behavior associations; changes in neural activity predicted more successful CP (psychometric slopes) for active but not passively evoked ERPs. Our results demonstrate auditory categorization is influenced by attention (active > passive) and is stronger for more familiar/overlearned stimulus domains (speech > music). In contrast to previous studies examining highly trained listeners (i.e., musicians), we infer that (i) CP skills are largely domain-specific and do not generalize to stimuli for which a listener has no immediate experience and (ii) categorical neural processing requires active engagement with the auditory stimulus.

Introduction

Human perception necessitates that the brain maps an infinite number of stimulus features into a smaller set of abstract groupings, an operation exemplified by categorical perception (CP). Original conceptions of CP assumed a 'specialized mode' of listening (Liberman & Mattingly, 1989) where categories only applied to speech. In perceptual studies, CP is indicated when gradually morphed speech sounds are heard as one of only a few discrete phonetic classes with an abrupt shift in perception near the midpoint of a stimulus continuum (Liberman *et al.*, 1967; Pisoni, 1973; Harnad, 1987; Pisoni & Luce, 1987). However, later studies revealed that CP is more general and that categories manifest in many aspects of human cognition including the perception of faces (Beale & Keil, 1995) and colors (Franklin *et al.*, 2008). Germane to audition, CP is also observable in the perceptual organization of music (Locke & Kellar,

1973; Siegel & Siegel, 1977; Burns & Ward, 1978; Zatorre & Halpern, 1979; Howard *et al.*, 1992; Burns & Campbell, 1994; Klein & Zatorre, 2011). Despite the ubiquitous role of speech and music to audition, we are aware of few studies directly contrasting auditory categorization between these domains behaviorally (cf. Cutting & Rosner, 1974; Weidema *et al.*, 2016) and none of which have compared their neural underpinnings. Such a comparison would help shed light on the domain-specificity of CP, its neural mechanisms, and how stimulus familiarity modulates this fundamental mode of hearing.

Several studies have indeed confirmed that CP is malleable to experience, learning, and stimulus familiarity. While categorical boundaries emerge early in life (Eimas *et al.*, 1971), they are further modified by one's native language experience (Kuhl *et al.*, 1992; Xu *et al.*, 2006b; Bidelman & Lee, 2015). Compared with non-native listeners, native speakers show larger and more categorical neural processing of speech sound contrasts from their native language (Zhang *et al.*, 2005; Xu *et al.*, 2006a; Bidelman & Lee,

Correspondence: Gavin M. Bidelman, as above.

E-mail: g.bidelman@memphis.edu

2015). Similarly, in the domain of music, trained musicians show sharper categorical boundaries for pitch intervals of the musical scale than their nonmusician peers (Burns & Ward, 1978; Zatorre, 1983; Burns & Campbell, 1994). Conceivably, such long-term experience with the acoustic signals of a certain domain (i.e., language, music) strengthens the representations of its sound identities and thus enhances categorical processing in a domain-specific manner. Domain-specificity in auditory CP is further evident by short-term perceptual learning studies, which generally fail to find robust retention and generalizability of categories after they are initially acquired (e.g., Lively *et al.*, 1993; Bradlow *et al.*, 1997). Collectively, these findings might be interpreted in relation to stimulus familiarity and/or uncertainty: less familiar sounds not encountered in a listener's regular experience fail to perceptually organize in a categorical manner. The first aim of the present study was to directly compare speech vs. musical sound classification in musically naïve listeners. In the light of previous work, we predicted that nonmusicians would show stronger categorical processing of speech compared with musical sounds given the overlearned nature and experience with the former and inexperience with the latter.

Our second aim was to evaluate the role of attention on auditory categorization. Goal-directed attention is known to amplify the auditory event-related potentials (ERPs) to complex sounds (Picton *et al.*, 1971; Hillyard *et al.*, 1973; Alain & Izenberg, 2003; Alain, 2007), affecting pre-perceptual processing as early as the N1-P2 complex, < 100–200 ms after the onset of sound (e.g., Schwent & Hillyard, 1975). However, while attention is known to modulate early auditory cortical processing, its role in CP has been equivocal.

Recent intracranial recordings in epileptic patients provocatively suggest that abstract speech categories arise within the superior temporal gyrus in the absence of goal-directed attention during passive listening (Chang *et al.*, 2010). In our own source analyses of the ERPs, we have shown that speech categories can emerge as early as primary auditory cortex under tasks in which listeners are actively attending to speech (Bidelman & Lee, 2015). However, in reconciling studies, it remains unclear if attentional engagement is necessary for the brain to form auditory categories (cf. Bidelman *et al.*, 2013b) or if they can emerge pre-attentively through automatic processes (cf. Chang *et al.*, 2010). The possibility of pre-attentive categorization is partially supported by computational modeling (Guenther & Gjaja, 1996) and limited animal recordings (Bizley & Cohen, 2013), which imply that sound representations in auditory cortex can self-organize because of non-uniformities in cell firing between exemplar vs. non-exemplar sounds (cf. within vs. between category tokens). Under this premise, the brain may partially categorize speech stimuli even prior to attentional engagement given its high degree of familiarity and overlearned nature.

To address the influences of attention and domain-specificity in auditory categorization, we recorded neuroelectric brain responses (ERPs) from nonmusician listeners as they rapidly categorized sounds along a comparable speech and music continua. Comparisons between music and speech allowed us to clarify the salience of CP for different classes of auditory stimuli and determine whether it is domain-specific to speech or if the process is generalized more broadly to other sound domains. Under the notion that speech is highly familiar and music unfamiliar to untrained listeners, we hypothesized that speech would evoke stronger CP and categorical coding than music, which was predicted to show weaker categorical processing. Comparisons between active and passive tasks using identical stimuli allowed us to examine whether or not categorical processing in the auditory system requires active attentional engagement (e.g., Bidelman *et al.*, 2013b; Alho *et al.*, 2016) or instead,

reflects automatic, pre-attentive grouping mechanisms (e.g., Joanisse *et al.*, 2007; Chang *et al.*, 2010).

Methods

Participants

Ten young adults (one male, nine females; age: $M = 22.5$, $SD = 2.8$ years) were recruited from the University of Memphis student body to participate in the experiment. All participants exhibited normal hearing sensitivity determined by an audiometric screening (i.e., < 25 dB HL between 500–2000 Hz). Each listener was strongly right-handed (78% laterality index; Oldfield, 1971) and had obtained a collegiate level of education. Age (Bidelman *et al.*, 2014a), tone-language experience (Bidelman & Lee, 2015), and musical training (Bidelman *et al.*, 2014b; Bidelman & Alain, 2015a) are known to modulate the early auditory cortical ERPs and CP. To avoid these confounds, individuals were excluded from participation if they reported any history of brain injury or psychiatric problems, were younger than 18 or older than 35 years, were familiar with a tonal language (e.g., Chinese), or had formal musical training on any combination of instruments totaling more than 3 years throughout their lifetime. All participants were paid for their time and gave informed consent in compliance with the Declaration of Helsinki and a protocol approved by the Institutional Review Board at the University of Memphis.

Stimuli

Speech continuum

We used a synthetic five-step vowel continuum to assess CP for speech (Bidelman *et al.*, 2013b, 2014b). Speech spectrograms are shown in the top row of Fig. 1. Each token of the continuum was separated by equidistant steps acoustically based on first formant frequency (F1), yet was perceived categorically from /u/ to /a/. Tokens were 100 ms, including 10 ms of rise/fall time to reduce spectral splatter in the stimuli. Each contained an identical voice fundamental (F0), second (F2), and third formant (F3) frequencies (F0: 150, F2: 1090, and F3: 2350 Hz). The F1 was parameterized 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.*, 2013b).

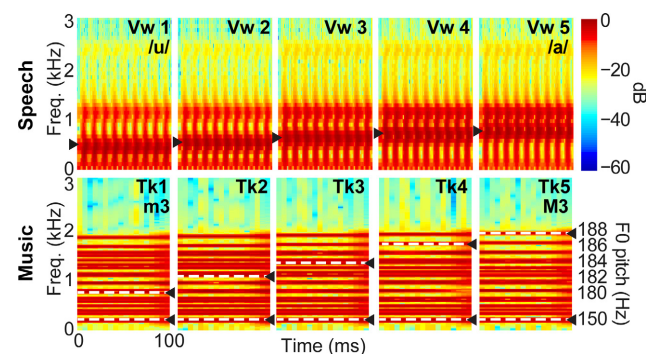


FIG. 1. Categorical speech (top) and music (bottom) continua spectrograms. In the speech continuum, first formant frequency was parameterized over five equal steps from 430 to 730 Hz (arrow) such that the resulting stimulus set spanned a perceptual phonetic continuum from the vowel /u/ to /a/. For music, complex tones were used to construct two-tone intervals (white dotted line) spanning a continuum from the minor (m3) to major (M3) third pitch interval along the chromatic musical scale. Both speech and music stimuli otherwise had identical duration (100 ms), intensity (83 dB SPL), and fundamental frequency (150 Hz). [Colour figure can be viewed at wileyonlinelibrary.com].

Music continuum

We synthesized a comparable five-step continuum of pitch intervals to assess the CP for musical sounds (Fig. 1, bottom). The stimulus set spanned five equidistant stimuli between a minor and major third interval on the chromatic scale (e.g., Burns & Ward, 1978). Individual notes were synthesized using complex tones consisting of 10 equal amplitude harmonics added in cosine phase. For each musical token, the lower of the two pitches was fixed with a F0 of 150 Hz (matching the F0 of the speech continuum) while the upper tone's F0 was varied over five equal steps to produce a perceptual continuum of musical intervals between the minor (m3; $f_{\text{lower}} = 150$, $f_{\text{higher}} = 180$ Hz) and major (M3; $f_{\text{lower}} = 150$, $f_{\text{higher}} = 188$ Hz) third. As with the speech continuum, music stimulus waveforms were 100 ms in duration including a 10 ms rise/fall time to reduce spectral splatter in the stimuli. The m3-M3 continuum was selected as these intervals occur most frequently in Western tonal music, are familiar even to nonmusicians (Brattico *et al.*, 2009), and evoked similar neural responses to trained musicians (Brattico *et al.*, 2009). Moreover, they connote the typical valence of 'sadness' (m3) and 'happiness' (M3) and are thus easily described to participants unfamiliar with music-theoretic labels.

Task and procedure

Listeners heard 200 trials of each individual speech/music token presented in either an active or passive listening condition spread over four blocks. Active blocks consisted of a speeded identification task, run separately for the speech and music continua. Passive blocks did not contain an overt task. For passive presentation, participants were instructed to ignore the sounds they hear and watch a self-selected movie with subtitles to maintain a calm and wakeful state (Bidelman *et al.*, 2014b). Active/passive and speech/music blocks were presented in alternating order.

Data acquisition and response evaluation were otherwise similar to previous reports from our laboratory (Bidelman *et al.*, 2013b, 2014b). Stimuli were delivered binaurally at an intensity of 83 dB SPL through insert earphones (ER-2; Etymotic Research). For active blocks, 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' for speech; 'm3/' or 'M3/' for music). The interstimulus interval (ISI) was jittered randomly between 400 and 600 ms (20 ms steps, rectangular distribution) following the listener's behavioral response which lasted ~ 500 ms (see Fig. 2B). The auditory cortical ERPs experience an attenuation at higher rates when stimuli are presented faster than 1 second intervals (Picton *et al.*, 1977; Bidelman, 2015c). Hence, for passive blocks, the ISI was jittered between 1150 and 1350 ms to ensure that the overall rate of stimulus delivery was comparable between active and passive conditions. Thus, with the 100 ms stimulus, active and passive stimulus onset asynchronies were each between about 1100–1450 ms, outside the typical rate-attenuation effects (Picton *et al.*, 1977).

EEG recording and preprocessing

EEGs were recorded from 64 sintered Ag/AgCl electrodes at standard 10–10 locations around the scalp (Oostenveld & Praamstra, 2001) using procedures described in our previous reports (Bidelman,

2015b; Bidelman & Lee, 2015). Continuous EEGs were digitized using a sampling rate of 500 Hz (SynAmps RT amplifiers; Compumedics Neuroscan) and an online passband of DC–200 Hz. Electrodes placed on the outer canthi of the eyes and the superior and inferior orbit were used to monitor ocular movements. During acquisition, electrodes were referenced to an additional sensor placed ~ 1 cm posterior to the Cz channel. Following data collection, data were re-referenced off-line to the common average reference. Contact impedances were maintained < 10 k Ω during data collection.

Subsequent preprocessing was performed in Curry 7 (Compumedics Neuroscan) and custom routines coded in MATLAB[®] 2015 (The MathWorks). Ocular artifacts (i.e., blinks and saccades) were corrected in the continuous EEG using a principal component analysis (PCA) (Wallstrom *et al.*, 2004). The PCA decomposition provided a set of components which best explained the topography of the blink/saccadic artifacts from a signal variance perspective. Following conventions in our lab (Bidelman & Howell, 2016) and the literature, the scalp projection of the first two PCA loadings was subtracted from the continuous EEG traces to nullify ocular contamination in the final ERPs. Cleaned EEGs were then digitally filtered (1–30 Hz; zero-phase filters), epoched (–200–800 ms, where $t = 0$ was stimulus onset), baselined to the pre-stimulus interval, and subsequently averaged in the time domain to obtain ERPs for each stimulus condition per participant. This resulted in 20 ERP waveforms per participant (five tokens*two listening conditions*two stimulus domains).

Data analysis

ERP data

To minimize potential bias and data reduction in electrode selection for analysis, we collapsed a subset of the 64-channel sensor data into a single region of interest (ROI) encompassing a cluster of six frontocentral electrodes (F1, Fz, F2, FC1, FCz, and FC2). This ROI was guided by our previous reports on the neural correlates of CP, which found that categorical effects in the auditory ERPs were most prominent at frontocentral scalp locations, indicative of bilateral sources proximal to the Sylvian fissure (Bidelman *et al.*, 2013b, 2014b; Bidelman & Lee, 2015).

We measured peak amplitude and latency for each of the prominent deflections of the cortical ERPs (P1, N1, P2). Prior work has shown that of the obligatory ERP components, the N1 and P2 waves are the most sensitive to speech perception tasks (Wood *et al.*, 1971; Alain *et al.*, 2007, 2010; Bidelman *et al.*, 2013b; Tremblay *et al.*, 2014) and the neuroplastic effects of speech-sound learning (Reinke *et al.*, 2003; Shahin *et al.*, 2003; Tremblay *et al.*, 2014). Furthermore, our previous studies suggested that the neural correlates of CP emerge around the time-frame of N1 and are fully manifested by P2 (Bidelman *et al.*, 2013b; Bidelman & Alain, 2015b; Bidelman & Lee, 2015). Consequently, we focused our primary analysis on the N1–P2 as we had specific hypotheses regarding how this overall signature would be modulated by auditory categorization. N1 was taken as the maximum negative-going trough between 70 and 115 ms and P2 as the positive-going peak between 140–180 ms, guided by the grand average across subjects. Response amplitudes for each condition was then computed as the peak-to-peak amplitude of the N1–P2 complex (e.g., Fig. 5). ERP analysis and automated peak selection was performed using custom routines coded in MATLAB.

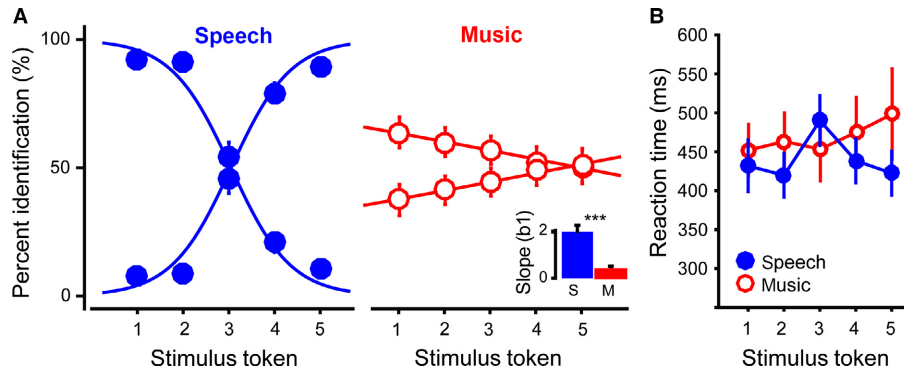


FIG. 2. Perceptual identification for speech and musical stimuli. (A) Psychometric identification functions show an abrupt shift in perception when classifying speech indicative of discrete perception (i.e., CP). In contrast, musical sounds are perceived in a continuous manner. (inset) Comparison of 'steepness' of identification functions (β_1 parameter). Sharper identification curves are observed for speech, indicating stronger CP than for music. (B) Reaction times for auditory identification. For speech, listeners are slower to label sounds near the categorical boundary (vw 3), indicative of categorical hearing (Pisoni & Tash, 1974; Bidelman *et al.*, 2013b). This modulation is not observed in labeling speeds for musical stimuli, indicating less salient CP in the music domain. Error bars = ± 1 SEM; *** $P < 0.0001$. [Colour figure can be viewed at wileyonlinelibrary.com].

To directly evaluate whether ERPs showed categorical coding, we averaged N1-P2 amplitudes to prototypical tokens at the end of the continua and compared this combination to the ambiguous token at its midpoint. This contrast [i.e., mean(Tk1, Tk5) vs. Tk 3] allowed us to assess the degree to which neural responses differentiated stimuli with well-formed categories from those heard with a bistable (ambiguous) identity within each domain. This measure is thought to reflect the degree of categorical neural coding indexed by the ERPs (e.g., Bidelman, 2015a).

Behavioral data

For each continuum, individual identification scores were fit with a two-parameter sigmoid function: $P = 1/[1 + e^{-\beta_1(x-\beta_0)}]$, where P is the proportion of trials identified as a given vowel, x is the step number along the stimulus continuum, and β_0 and β_1 the location and slope of the logistic fit estimated using nonlinear least-squares regression. Comparing parameters between conditions revealed possible differences in the location and 'steepness' (i.e., rate of change) of the categorical boundary as a function of stimulus domain (i.e., speech vs. music). Larger β_1 values reflect steeper psychometric functions and hence, indicate stronger CP in one stimulus domain vs. the other.

Behavioral speech labeling speeds [i.e., reaction times (RTs)] were computed as listeners' median response latency across trials for a given condition. RTs outside 250–2500 ms were deemed outliers and excluded from further analysis (Bidelman *et al.*, 2013b). One participant's behavioral data was not logged in the music condition because of hardware malfunction. Consequently, to allow for regression analysis on the full data complement, this missing value was replaced by the mean of the remaining cohort.

Statistical analysis

Unless otherwise noted, all dependent measures were analyzed separately by domain (speech vs. music) using two-way, mixed model ANOVAs (subject = random factor) with fixed effects of listening condition (two levels: active vs. passive) and token type [two levels: prototypical (Tk1/5) vs. ambiguous (Tk3) tokens] (PROC GLIMMIX, SAS[®] 9.4; SAS Institute, Inc.). Tukey–Kramer multiple comparisons controlled Type I error inflation. An *a priori* significance level was set at $\alpha = 0.05$.

Brain-behavior relations

To examine the degree to which neural representations for speech and music stimuli predicted behavioral categorization, we performed weighted least square regression between listeners' neural responses and their β_1 values (i.e., steepness of their psychometric function). Robust fitting was achieved using 'fitlm' in MATLAB. No influential outliers in the data were identified in initial diagnostics before conducting regression (DFFITS test; Montgomery, 2005). Psychometric slopes (β_1) reflect the degree to which listeners distinguish prototypical from ambiguous categories and thus the salience of CP. To arrive at a comparable and single measure to describe how neurophysiological responses distinguish prototypical from ambiguous sounds (i.e., show categorical coding), we derived a new variable from listeners' ERPs (Δ ERP magnitude), computed as the difference between their N1-P2 amplitudes evoked by the Tk1/5 tokens and the ambiguous case (Tk3) (i.e., mean(Tk1/5) – Tk3; Bidelman, 2015a). We then regressed Δ ERP values against behavioral β_1 responses. This regression analysis assessed the degree to which changes in neural activity to speech/music reflect their successful behavioral categorization. Studies have suggested that both actively- (Bidelman *et al.*, 2013b) and passively evoked (Chang *et al.*, 2010) speech activity might be correlated with listeners' behavioral psychometrics for CP. We reasoned that for a neural (active or passive) response to truly reflect behavior, it should be correlated with perceptual measures.

Results

Behavioral identification (% and RTs)

Behavioral psychometric identification functions and RTs are shown for the speech and music continua in Fig. 2A and B, respectively. Listeners' identification was more dichotomous for speech compared with music stimuli, as indicated by an abrupt shift in the perception midway through the continuum of the former. In contrast, music elicited largely continuous perception as indicated by the lack of any abrupt perceptual shift. This was confirmed by an independent samples *t*-test (two-tailed) conducted on psychometric identification slopes (β_1) (Fig. 2A, inset), which revealed steeper identification when identifying speech compared to music ($t_9 = 7.88$, $P < 0.0001$). These findings suggest that in nonmusicians,

perceptual categorization is stronger when classifying speech compared with musical sounds.

Behavioral RTs for auditory classification in the speech and music domain are shown in Fig. 2B. Speech RTs showed a strong modulation across the continuum ($F_{4,36} = 10.78$, $P < 0.0001$) whereas music stimuli did not ($F_{4,36} = 1.86$, $P = 0.14$). For speech, the slowing of classification speeds for tokens near the CP boundary was confirmed by an *a priori* test contrasting RTs on (vw3 vs. mean(vw1/2/4/5): $F_{4,36} = 40.19$, $P < 0.0001$). A slowing in speech labeling near the perceptual boundary is an additional hallmark of CP observed in previous behavioral studies (Pisoni & Tash, 1974; Bidelman *et al.*, 2013b, 2014b). The fact that this effect is not observed for music further indicates that our listeners heard speech more categorically than musical sounds.

Cortical ERPs

Grand average ERP waveforms and scalp topographies for each stimulus domain (speech vs. music) and listening task (active vs. passive) are shown in Figs 3 and 4, respectively. Topographies show the distribution of the N1-P2 complex, the primary neural measure of the study (see Fig. S1, Supporting information for topographies of the individual N1 and P2 waves). Scalp maps confirm that evoked responses to speech and music stimuli were maximal over frontocentral electrodes, consistent with neural generators in the supratemporal plane (Picton *et al.*, 1999). ERP waveforms extracted from the frontocentral electrode ROI (channels F1, Fz, F2, FC1, FCz, and FC2) revealed distinct modulations in the early ERPs (i.e., < 200 ms) within the timeframe of the N1-P2 complex between prototypical (Tk1/5) and ambiguous (Tk3) tokens. The strength of this modulation was highly dependent on whether stimuli were from the speech or music domain and listeners' attentional state.

Individual N1 amplitudes were not modulated with changes in listening condition within either the speech ($F_{1,9} = 1.75$, $P = 0.22$) or music domain ($F_{1,9} = 0.00$, $P = 0.99$). Similar results were observed for the P2 (speech: $F_{1,9} = 1.97$, $P = 0.19$; music: $F_{1,9} = 2.31$, $P = 0.94$). However, this might be expected in light of previous studies that have also observed weak effects in individual ERP waves during similar CP tasks (Bidelman *et al.*, 2014b) and the notion that CP emerges between the N1 and P2 components (Bidelman *et al.*, 2013b). Hence, we focused subsequent analysis on the N1-P2 complex as a whole.

Overall amplitude of the N1-P2 response was modulated by task manipulations (Fig. 5). For speech, an ANOVA conducted on N1-P2 amplitudes revealed a significant interaction between listening condition and stimulus token ($F_{1,9} = 5.97$, $P = 0.037$). Multiple comparisons revealed an attentional effect for the ambiguous speech (Tk 3), whereby ERPs were stronger in passive compared with active listening. Weaker responses in the active condition for Tk3 may reflect increased inhibition or top-down gating on sensory coding with the increased listening effort for perceptually ambiguous speech (Knight *et al.*, 1999; Bidelman *et al.*, 2014a). In contrast, prototypical vowels (Tk1/5) evoked similar N1-P2 amplitudes regardless of attentional state. Comparisons by listening condition revealed that N1-P2 amplitudes were larger in response to Tk1/5 than Tk3, indicative of categorical coding. This indicates that brain activity differentiated prototypical from ambiguous speech sounds (i.e., Tk1/5 > Tk3) producing stronger categorical coding for the former. In contrast, ERPs did not show categorical coding during passive listening (i.e., Tk1/5 = Tk3).

In contrast to speech, we found no significant effects of attentional state ($F_{1,9} = 0.01$, $P = 0.93$), stimulus token ($F_{1,9} = 0.00$, $P = 0.99$), nor their interaction ($F_{1,9} = 0.40$, $P = 0.54$) on neural responses to music stimuli. Collectively, these findings corroborate the behavioral data by revealing that the neural correlates of CP (at least in nonmusicians) are stronger for speech compared to musical sounds and are only evident during active listening conditions.

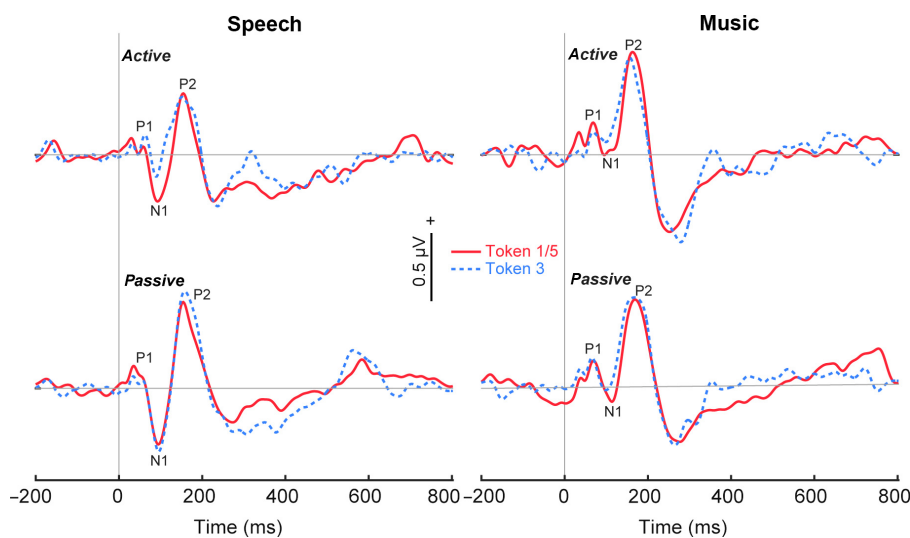


FIG. 3. Cortical ERP waveforms for speech and musical stimuli under active and passive listening. Traces denote the average potential for each condition extracted from a frontocentral cluster of electrodes (F1, Fz, F2, FC1, FCz, FC2). Gray lines denote the onset of the time-locking stimulus at $t = 0$. Prominent deflections of the evoked response (P1, N1, and P2) are demarcated along the time courses. Waveforms are shown contrasting prototypical and ambiguous sounds along each continuum. Differences between traces [mean (Tk1 and Tk5)] > Tk 3) are indicative of categorical coding. Modulations in response morphology appear in the timeframe of the N1-P2 complex indicating categorical neural coding emerging ~ 100–200 ms after stimulus onset. [Colour figure can be viewed at wileyonlinelibrary.com].

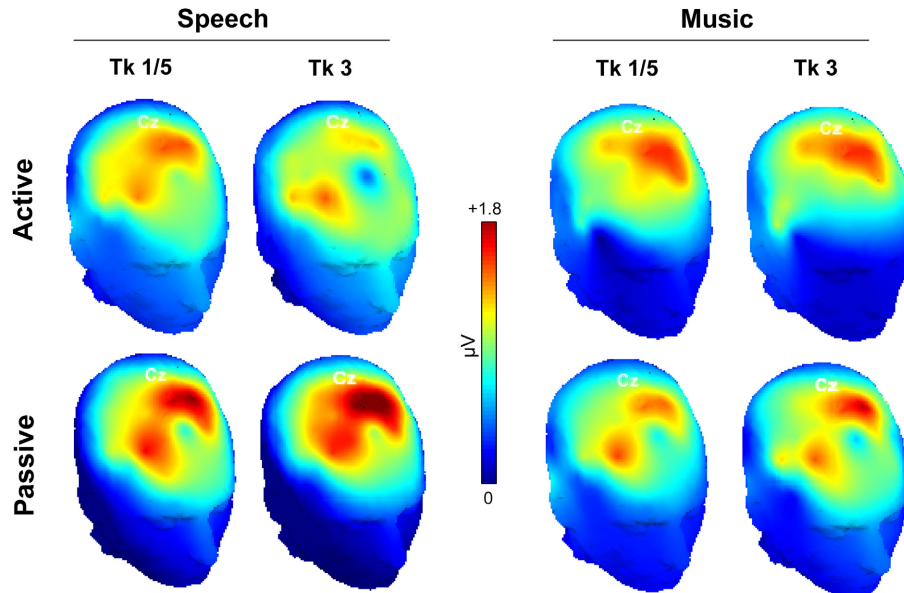


FIG. 4. ERP scalp topographies for speech and musical tokens. Topographies show the distribution of the N1-P2 response (see Fig. S1 for individual topographies of the N1 and P2 waves). For both stimulus domains, evoked responses are distributed maximally over frontocentral regions of the scalp. Larger categorical differentiation (Tk1/5 > Tk3) is observed for actively attended speech, particularly near frontal midline electrodes (Fz). [Colour figure can be viewed at [wileyonlinelibrary.com](#)].

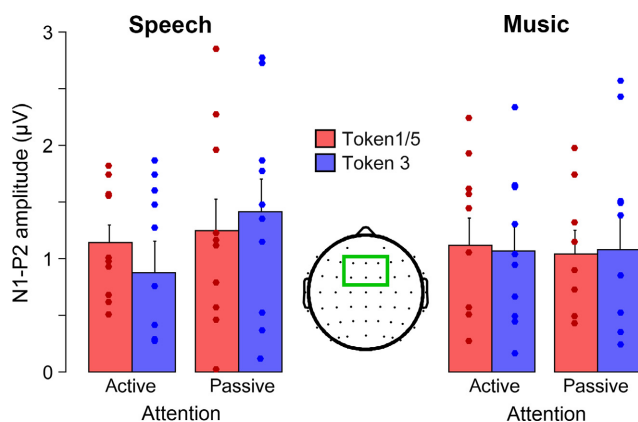


FIG. 5. Effects of attentional load and stimulus domain on the categorical encoding of speech and music stimuli. (*inset*) Frontocentral electrode ROI used to quantify N1-P2 amplitudes. For speech, prototypical vowels (Tk1/5) result in stronger categorical processing compared with the ambiguous tokens (vw3) during active speech listening. No attentional or categorical differentiation is observed under passive listening or for music stimuli. Individual data points represent single participants. Error bars = ± 1 SEM. [Colour figure can be viewed at [wileyonlinelibrary.com](#)].

Brain-behavioral relations

Regression was used to assess the degree to which changes in neural activity between prototypical vs. ambiguous sounds [i.e., (Tk1/5) – Tk3] reflect their successful behavioral categorization (Fig. 6). Results revealed that the steepness of listeners' psychometric functions was positively associated with Δ ERP responses such that larger changes in neural activity predicted stronger behavioral CP (Fig. 6A) ($R^2 = 0.30$, $t_{18} = 2.77$, $P = 0.0127$). However, separate analyses by domain indicated this effect was driven by a relation between neural and behavioral responses only for speech ($R^2 = 0.42$, $t_8 = 2.38$, $P = 0.0446$); no brain-behavior relation was observed when considering music stimuli alone ($R^2 = 0.01$, $t_8 = -0.07$, $P = 0.95$). As a further confirmation of these findings,

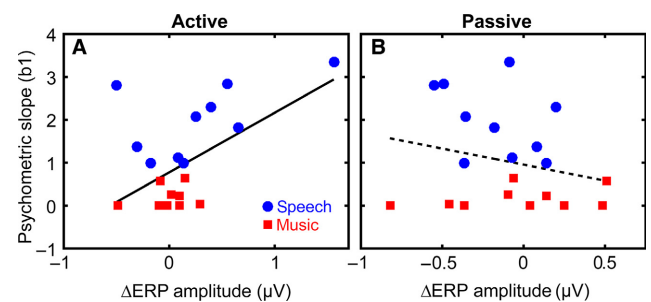


FIG. 6. Brain-behavior relations reveal that perceptual categorization is dictated by auditory neural encoding during active but not passive listening. Shown here are the regressions between the change in N1-P2 amplitude between prototypical vs. ambiguous sounds [i.e., Δ ERP = (Tk1/5) – Tk3] and the slope of listeners' psychometric identification functions (β_1 parameter). A larger Δ ERP differential, reflecting stronger categorical coding, is associated with steeper, more dichotomous perceptual CP during active listening (solid line; $P = 0.0127$). (B) Brain-behavior associations are not observed under passive listening (dotted line; $P \gg 0.05$) suggesting that successful CP requires active attentional deployment. [Colour figure can be viewed at [wileyonlinelibrary.com](#)].

we similarly regressed Δ ERP responses elicited in the passive listening task against behavioral psychometric slopes as a control analysis. We reasoned that if the Δ ERP measure truly reflected behavior (as observed for active responses), passive responses should *not* be associated with behavioral measures given there was no task during their elicitation. Some studies have suggested that passively evoked speech activity is highly predictive of listeners' behavioral CP (Chang *et al.*, 2010). This control analysis confirmed there was no association between passively evoked amplitudes and behavioral CP for either speech ($R^2 = 0.13$, $t_8 = -1.10$, $P = 0.30$) or musical stimuli ($R^2 = 0.10$, $t_8 = 0.93$, $P = 0.37$) (Fig. 6B). These findings confirm the degree to which brain activity (during active listening) distinguishes prototypical from ambiguous tokens predicts the behavioral salience of CP experienced by listeners (cf. Bidelman, 2015a).

Discussion

By measuring electrical brain activity elicited by categorically perceived speech and musical stimuli, results of the current study relate to three main findings: (i) complex sounds are processed according to their categorical membership (in addition to their acoustic properties) within ~ 200 ms following sound onset, (ii) speech sounds are more categorically organized than musical stimuli (in nonmusician listeners), (iii) auditory categorization manifests only under active attentional engagement and does not occur automatically under passive sound exposure. Our findings bolster the notion that CP can occur within the first few hundred milliseconds of the auditory cortical hierarchy (Chang *et al.*, 2010; Bidelman *et al.*, 2013b, 2014b) and further extend previous studies on CP by demonstrating its neural substrates are modulated by the degree of stimulus familiarity (i.e., domain-specificity) and top-down mechanisms (i.e., attention).

Domain-specificity reveals the influence of stimulus familiarity on auditory categorization

We found that listeners showed stronger CP for speech than music as evident by their sharper psychometric identification and faster speed in classification decisions. These behavioral data were accompanied by biological enhancements in the neural differentiation of prototypical vs. ambiguous speech sounds and stronger link to behavioral under speech compared with music listening. More prominent CP for speech rather than musical sounds is consistent with the notion that stimulus familiarity modulates categorical representations. Indeed, we have previously shown that familiar speech contexts that are more frequently encountered in daily communication elicit stronger categorical responses in regions adjacent to primary auditory cortex compared to unfamiliar contexts (Bidelman & Lee, 2015). Our findings are also consistent with results reported in infant (Kuhl *et al.*, 1992; McCandliss *et al.*, 2002; McClelland *et al.*, 2002) and cross-language studies (Werker & Tees, 1984; Zhang *et al.*, 2005; Kirmse *et al.*, 2008; Bomba *et al.*, 2011) which demonstrate that listeners are sensitized to the linguistic cues utilized more frequently within their native language and become desensitized/prune auditory cues not readily necessary for their daily communication (Kuhl *et al.*, 1992). Similarly, musical training has been shown to increase the identification accuracy for musical intervals relative to nonmusicians further confirming domain-specific tuning of CP (Cutting & Rosner, 1974; Burns & Ward, 1978). Although these studies utilized different populations (e.g., infants, non-native speakers, musicians), results of our current experiment broadly support the notion that an individual's familiarity with a particular auditory domain influences their categorization of sounds within that domain.

Under this interpretation, listeners are likely to become more sensitized to sounds of their domain and more readily establish categorical identities for those auditory signals. This explanation may account for the stronger CP we find for speech compared with music. Our cohort of listeners consisted of English-speaking nonmusicians. Consequently, they had long-term experience with the vowel speech stimuli of our study but lacked formal labels for the musical sounds. Indeed, we observed that prototypical speech sounds elicited strong CP in both neural and behavioral responses compared with ambiguous sounds. Sharper CP in speech suggests that the salience of categorical representations in this auditory domain were stronger, perhaps given their higher familiarity. Participants in our sample had minimal formal musical training and

less familiarity with our musical stimuli resulting in weaker CP compared to speech. Stronger categorical organization in one domain could be realized neurobiologically via reorganization of its sensory representations; given its overlearned nature, speech stimuli might act to warp or restrict the perceptual space near category boundaries to supply a more dichotomous decision when classifying sound objects.

Our results also demonstrate that CP skills do not generalize broadly to other sounds. Despite robust CP and neural coding for speech, listeners in our cohort showed continuous perception in the music domain (e.g., Fig. 2A). This suggests that auditory categorization skills are largely domain-specific and do not necessarily transfer to complementary auditory modalities. Notably, these findings differ from what is observed in highly trained listeners (Bidelman *et al.*, 2014b; Wu *et al.*, 2015; Chang *et al.*, 2016; Weidema *et al.*, 2016). Musicians, for example, show benefits in speech categorization, indicating that their specific experience with manipulating musical pitch can transfer (generalize) to enhance categorization in the speech domain (Cooper & Wang, 2012; Bidelman *et al.*, 2014b; Bidelman & Alain, 2015b). This asymmetry in perceptual transfer is reminiscent of other prominent asymmetries observed between speech and music processing (e.g., McMurray *et al.*, 2008; Bidelman *et al.*, 2013a; Weidema *et al.*, 2016) and may reflect the fact that music places a higher demand on the precision of auditory processing than speech (Bidelman *et al.*, 2011, 2013a; Patel, 2011; Hutka *et al.*, 2015).

Time course of CP

Our ERP data reveal robust neural correlates of CP within the time frame of the N1-P2 deflection, roughly 150–200 ms after sound arrives at the ear. Coupled with our previous studies, data here imply that the phonetic (categorical) representations of speech emerge before 150 ms (Bidelman *et al.*, 2013b; Bidelman & Alain, 2015b), and possibly as early as primary auditory cortex under some circumstances (e.g., highly experienced listeners: Bidelman & Lee, 2015). ERP topographies (Fig. 4) and the polarity reversal between frontocentral scalp regions and the mastoids are consistent with generators in the auditory cortices along the Sylvian fissure (Picton *et al.*, 1999). We have not observed categorical representations lower than primary auditory cortex (e.g., brainstem) (Bidelman *et al.*, 2013b), suggesting that these early cerebral structures are good initial candidates for a categorical processor. It is important to clarify that our data do not diminish the involvement of other important brain regions downstream from the lemniscal auditory areas in categorical processing (e.g., pars opercularis and prefrontal regions: Myers *et al.*, 2009; Alho *et al.*, 2016). Indeed, top-down influences could partly account for the categorical modulations we observe in the auditory N1-P2 (for early prefrontal modulation of auditory sensory responses, see Knight *et al.*, 1980; Knight, 1994). While our data cannot rule out this possibility, we note that such feedback effects would need to occur very early (100–150 ms) and while the stimulus is still being encoded to account for our findings. Nevertheless, our results are consistent with previous ERP studies (Dehaene-Lambertz, 1997; Phillips *et al.*, 2000; Bidelman *et al.*, 2013b, 2014a,b; Altmann *et al.*, 2014; Bidelman & Lee, 2015), fMRI data (Binder *et al.*, 2004; Kilian-Hütten *et al.*, 2011), and near-field cellular recordings (Steinschneider *et al.*, 2003; Micheyl *et al.*, 2005; Bar-Yosef & Nelken, 2007; Chang *et al.*, 2010) which suggest that auditory cortical areas code more than low-level acoustic features; instead they implicate these early regions in the formation of

perceptual representations, auditory objects, and abstract sound categories.

The role of attention in auditory CP

Comparisons between actively and passively evoked ERPs to identical auditory stimuli allowed us to assess the role of attention on auditory categorization. Previous studies have been equivocal on this matter. Several studies have shown that attentional engagement is necessary for the brain to form categories (Bidelman *et al.*, 2013b; Alho *et al.*, 2016) while others have implied that categorical representations might emerge pre-attentively through automatic processes (Joanisse *et al.*, 2007; Chang *et al.*, 2010). However, in most previous studies (Joanisse *et al.*, 2007; Chang *et al.*, 2010; Bidelman *et al.*, 2013b), attention was not directly manipulated since either an active *or* passive task was used, but not both. The current investigation therefore helps clarify the relationship between attention and auditory categorical brain processing. Converging with previous MEG studies directly contrasting active and passive speech listening tasks (Alho *et al.*, 2016), our results provide evidence that auditory categorization likely manifests under active attentional engagement rather than through automatic, passive sound exposure (but see Joanisse *et al.*, 2007; Chang *et al.*, 2010). This is supported by the fact that ERPs showed categorical organization (i.e., prototypical > ambiguous speech) (Fig. 5) and were predictive of behavioral CP only in active (but not passive) conditions (cf. Fig. 6A vs. B).

That attention is necessary to form auditory categories is supported by recent E/MEG neuroimaging studies (Bidelman *et al.*, 2013b; Alho *et al.*, 2016). However, our findings also contrast previous studies suggesting that categorical representations might occur pre-attentively without an overt response or a conscious perceptual task (cf. Joanisse *et al.*, 2007; Chang *et al.*, 2010). Differences in neuroimaging techniques offer one potential explanation for differences across studies. In particular, it is possible that detailed near-field recordings which provide higher spatial resolution than EEG may be more helpful in revealing categorical-like coding pre-attentively (Chang *et al.*, 2010). Alternatively, aspects of the task, including the possibility of covert listening, may account for some of the discrepancies. For example, in studies where attention is not directly manipulated nor controlled (e.g., Joanisse *et al.*, 2007; Chang *et al.*, 2010), it is possible that electrophysiological measures could be partially influenced by participants inadvertently shifting attention while listening to speech sounds. Direct comparisons between active and passive speech tasks using the same stimuli (this study; Alho *et al.*, 2016) are needed to fully resolve attentional effects on CP. Nevertheless, under this design, we fail to find strong evidence for neural correlates of CP under strictly passive listening.

Conclusions

In sum, we infer that the neural underpinnings of CP and robust behavioral identification require active task engagement with stronger categorical neural coding expressed only during states of attentional deployment (cf. Chang *et al.*, 2010). Further, the fact that CP is stronger for speech than music (at least in nonmusician listeners) indicates that the formation of categories is contingent, at least in part, upon stimulus familiarity. Lastly, we infer that skilled categorization is largely domain-specific and does not necessarily generalize to stimuli for which a listener has no immediate experience unlike results observed in highly trained listeners (e.g., Bidelman *et al.*, 2014b).

Supporting Information

Additional supporting information can be found in the online version of this article:

Fig. S1. Scalp topographies of the individual N1 (A) and P2 (B) waves.

Audio S1. audio file of the 5-step *music* stimulus continuum

Audio S2. audio file of the 5-step *speech* stimulus continuum

Acknowledgment

Portions of this work were supported by a grant from the GRAMMY Foundation® awarded to G.M.B.

Author contributions

GMB designed the experiment, BW collected the data, GMB and BW analyzed the data and wrote the paper.

References

- Alain, C. (2007) Breaking the wave: effects of attention and learning on concurrent sound perception. *Hearing Res.*, **229**, 225–236.
- Alain, C. & Izenberg, A. (2003) Effects of attentional load on auditory scene analysis. *J. Cognitive Neurosci.*, **15**, 1063–1073.
- Alain, C., Snyder, J.S., He, Y. & Reinke, K.S. (2007) Changes in auditory cortex parallel rapid perceptual learning. *Cereb. Cortex*, **17**, 1074–1084.
- Alain, C., Campeanu, S. & Tremblay, K.L. (2010) Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *J. Cognitive Neurosci.*, **22**, 392–403.
- Alho, J., Green, B.M., May, P.J.C., Sams, M., Tiitinen, H., Rauschecker, J.P. & Jääskeläinen, I.P. (2016) Early-latency categorical speech sound representations in the left inferior frontal gyrus. *NeuroImage*, **129**, 214–223.
- 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**, 13–23.
- Bar-Yosef, O. & Nelken, I. (2007) The effects of background noise on the neural responses to natural sounds in cat primary auditory cortex. *Front. Comput. Neurosci.*, **1**, 1–14.
- Beale, J.M. & Keil, F.C. (1995) Categorical effects in the perception of faces. *Cognition*, **57**, 217–239.
- Bidelman, G.M. (2015a) Induced neural beta oscillations predict categorical speech perception abilities. *Brain Lang.*, **141**, 62–69.
- Bidelman, G.M. (2015b) Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. *Hearing Res.*, **323**, 68–80.
- Bidelman, G.M. (2015c) Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. *J. Neurosci. Meth.*, **241**, 94–100.
- Bidelman, G.M. & Alain, C. (2015a) Hierarchical neurocomputations underlying concurrent sound segregation: connecting periphery to percept. *Neuropsychologia*, **68**, 38–50.
- Bidelman, G.M. & Alain, C. (2015b) Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J. Neurosci.*, **35**, 1240–1249.
- Bidelman, G.M. & Howell, M. (2016) Functional changes in inter- and intra-hemispheric auditory cortical processing underlying degraded speech perception. *NeuroImage*, **124**, 581–590.
- Bidelman, G.M. & Lee, C.-C. (2015) Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *NeuroImage*, **120**, 191–200.
- Bidelman, G.M., Gandour, J.T. & Krishnan, A. (2011) Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain Cognition*, **77**, 1–10.
- Bidelman, G.M., Hutka, S. & Moreno, S. (2013a) Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PLoS One*, **8**, e60676.

- Bidelman, G.M., Moreno, S. & Alain, C. (2013b) Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, **79**, 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. *Neurobiol. Aging*, **35**, 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. *Eur. J. Neurosci.*, **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. *Nat. Neurosci.*, **7**, 295–301.
- Bizley, J.K. & Cohen, Y.E. (2013) The what, where and how of auditory-perception. *Nat. Rev. Neurosci.*, **14**, 693–707.
- Bomba, M.D., Cholby, D. & Pang, E.W. (2011) Phoneme discrimination and mismatch negativity in English and Japanese speakers. *NeuroReport*, **22**, 479–483.
- Bradlow, A.R., Pisoni, D.B., Yamada, R.A. & Tohkura, Y. (1997) Training the Japanese listener to identify English /t/ and /l/. IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.*, **101**, 2299–2310.
- Brattico, E., Pallesen, K.J., Varyagina, O., Bailey, C., Anourova, I., Jarvenpaa, M., Eerola, T. & Tervaniemi, M. (2009) Neural discrimination of nonprototypical chords in music experts and laymen: An MEG study. *J. Cognitive Neurosci.*, **21**, 2230–2244.
- Burns, E.M. & Campbell, S.L. (1994) Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: examples of categorical perception. *J. Acoust. Soc. Am.*, **96**, 2704–2719.
- Burns, E.M. & Ward, W.D. (1978) Categorical perception – phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *J. Acoust. Soc. Am.*, **63**, 456–468.
- 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. *Nat. Neurosci.*, **13**, 1428–1432.
- Chang, D., Hedberg, N. & Wang, Y. (2016) Effects of musical and linguistic experience on categorization of lexical and melodic tones. *J. Acoust. Soc. Am.*, **139**, 2432–2447.
- Cooper, A. & Wang, Y. (2012) The influence of linguistic and musical experience on Cantonese word learning. *J. Acoust. Soc. Am.*, **131**, 4756–4769.
- Cutting, J.E. & Rosner, B.S. (1974) Categories and boundaries in speech and music. *Percept. Psychophys.*, **16**, 564–570.
- Dehaene-Lambertz, G. (1997) Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, **8**, 919–924.
- Eimas, P.D., Siqueland, E.R., Jusczyk, P. & Vigorito, J. (1971) Speech perception in infants. *Science*, **171**, 303–306.
- Franklin, A., Drivonikou, G.V., Clifford, A., Kay, P., Regier, T. & Davies, I.R. (2008) Lateralization of categorical perception of color changes with color term acquisition. *Proc. Natl. Acad. Sci. USA*, **105**, 18221–18225.
- Guenther, F.H. & Gjaja, M.N. (1996) The perceptual magnet effect as an emergent property of neural map formation. *J. Acoust. Soc. Am.*, **100**, 1111–1121.
- Harnad, S.R. (1987). *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, New York.
- Hillyard, S.A., Hink, R.F., Schwent, V.L. & Picton, T.W. (1973) Electrical signs of selective attention in the human brain. *Science*, **182**, 177–180.
- Howard, D., Rosen, S. & Broad, V. (1992) Major/Minor triad identification and discrimination by musically trained and untrained listeners. *Music Percept.*, **10**, 205–220.
- Hutka, S., Bidelman, G.M. & Moreno, S. (2015) Pitch expertise is not created equal: cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. *Neuropsychologia*, **71**, 52–63.
- Joannisse, M.F., Zevin, J.D. & McCandliss, B.D. (2007) Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cereb. Cortex*, **17**, 2084–2093.
- Kilian-Hütten, N., Valente, G., Vroomen, J. & Formisano, E. (2011) Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J. Neurosci.*, **31**, 1715–1720.
- Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schroger, E. & Jacobsen, T. (2008) Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *Int. J. Psychophysiol.*, **67**, 131–143.
- Klein, M.E. & Zatorre, R.J. (2011) A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia*, **49**, 878–887.
- Knight, R.T. (1994). Attention regulation and human prefrontal cortex. In Thierry, A.-M. (Ed.), *Motor and Cognitive Functions of the Prefrontal Cortex*. Springer-Verlag, Berlin, pp. 160–173.
- Knight, R.T., Hillyard, S.A., Woods, D.L. & Neville, H.J. (1980) The effects of frontal and temporal-parietal lesions on the auditory evoked potential in man. *Clin. Neurophysiol.*, **50**, 112–124.
- Knight, R.T., Staines, W.R., Swick, D. & Chao, L.L. (1999) Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta Psychol.*, **101**, 159–178.
- Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N. & Lindblom, B. (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255**, 606–608.
- Lieberman, A.M. & Mattingly, I.G. (1989) A specialization for speech perception. *Science*, **243**, 489–494.
- Lieberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychol. Rev.*, **74**, 431–461.
- Lively, S.E., Logan, J.S. & Pisoni, D.B. (1993) Training Japanese listeners to identify English /t/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *J. Acoust. Soc. Am.*, **94**, 1242–1255.
- Locke, S. & Kellar, L. (1973) Categorical perception in a non-linguistic mode. *Cortex*, **9**, 355–369.
- McCandliss, B.D., Fiez, J.A., Protopapas, A., Conway, M. & McClelland, J.L. (2002) Success and failure in teaching the [r]-[l] contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cogn. Affect. Behav. Ne.*, **2**, 89–108.
- McClelland, J.L., Fiez, J.A. & McCandliss, B.D. (2002) Teaching the /t/-/l/ discrimination to Japanese adults: behavioral and neural aspects. *Physiol. Behav.*, **77**, 657–662.
- McMurray, B., Dennhardt, J.L. & Struck-Marcell, A. (2008) Context effects on musical chord categorization: different forms of top-down feedback in speech and music? *Cognitive Sci.*, **32**, 893–920.
- Micheyl, C., Tian, B., Carlyon, R.P. & Rauschecker, J.P. (2005) Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*, **48**, 139–148.
- Montgomery, D.C. (2005). *Design and Analysis of Experiments*. John Wiley & Sons, Hoboken, NJ.
- Myers, E.B., Blumstein, S.E., Walsh, E. & Eliassen, J. (2009) Inferior frontal regions underlie the perception of phonetic category invariance. *Psychol. Sci.*, **20**, 895–903.
- Oldfield, R.C. (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, **9**, 97–113.
- Oostenveld, R. & Praamstra, P. (2001) The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.*, **112**, 713–719.
- Patel, A.D. (2011) Why would musical training benefit the neural encoding of speech? The OPERA hypothesis *Front. Psychol.*, **2**, 142.
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M. & Roberts, T. (2000) Auditory cortex accesses phonological categories: an MEG mismatch study. *J. Cognitive Neurosci.*, **12**, 1038–1055.
- Picton, T.W., Hillyard, S.A., Galambos, R. & Schiff, M. (1971) Human auditory attention: a central or peripheral process? *Science*, **173**, 351–353.
- Picton, T.W., Woods, D.L., Baribaeu-Braun, J. & Healy, T.M.G. (1977) Evoked potential audiometry. *J. Otolaryngol.*, **6**, 90–119.
- Picton, T.W., Alain, C., Woods, D.L., John, M.S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J. & Trujillo, N.J. (1999) Intracerebral sources of human auditory-evoked potentials. *Audiol. Neurootol.*, **4**, 64–79.
- Pisoni, D.B. (1973) Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept. Psychophys.*, **13**, 253–260.
- Pisoni, D.B. & Luce, P.A. (1987) Acoustic-phonetic representations in word recognition. *Cognition*, **25**, 21–52.
- Pisoni, D.B. & Tash, J. (1974) Reaction times to comparisons within and across phonetic categories. *Percept. Psychophys.*, **15**, 285–290.
- Reinke, K., He, Y., Wang, C. & Alain, C. (2003) Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Res.*, **17**, 781–791.
- Schwent, V.L. & Hillyard, S.A. (1975) Evoked potential correlates of selective attention with multi-channel auditory inputs. *Electroen. Clin. Neuro.*, **28**, 131–138.
- Shahin, A., Bosnyak, D.J., Trainor, L.J. & Roberts, L.E. (2003) Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J. Neurosci.*, **23**, 5545–5552.

- Siegel, J.A. & Siegel, W. (1977) Absolute identification of notes and intervals by musicians. *Percept. Psychophys.*, **21**, 143–152.
- 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. *J. Acoust. Soc. Am.*, **114**, 307–321.
- Tremblay, K.L., Ross, B., Inoue, K., McClannahan, K. & Collet, G. (2014) Is the auditory evoked P2 response a biomarker of learning? *Front. Syst. Neurosci.*, **8**, 1–13.
- Wallstrom, G.L., Kass, R.E., Miller, A., Cohn, J.F. & Fox, N.A. (2004) Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods. *Int. J. Psychophysiol.*, **53**, 105–119.
- Weidema, J.L., Roncaglia-Denissen, M.P. & Honing, H. (2016) Top-down modulation on the perception and categorization of identical pitch contours in speech and music. *Front. Psychol.*, **7**, 817.
- Werker, J.F. & Tees, R.C. (1984) Phonemic and phonetic factors in adult cross-language speech perception. *J. Acoust. Soc. Am.*, **75**, 1866–1878.
- Wood, C.C., Goff, W.R. & Day, R.S. (1971) Auditory evoked potentials during speech perception. *Science*, **173**, 1248–1251.
- Wu, H., Ma, X., Zhang, L., Liu, Y., Zhang, Y. & Shu, H. (2015) Musical experience modulates categorical perception of lexical tones in native Chinese speakers. *Front. Psychol.*, **6**, 1–7.
- Xu, Y., Gandour, J., Talavage, T., Wong, D., Dziedzic, M., Tong, Y., Li, X. & Lowe, M. (2006a) Activation of the left planum temporale in pitch processing is shaped by language experience. *Hum. Brain Mapp.*, **27**, 173–183.
- Xu, Y., Gandour, J.T. & Francis, A. (2006b) Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J. Acoust. Soc. Am.*, **120**, 1063–1074.
- Zatorre, R.J. (1983) Category-boundary effects and speeded sorting with a harmonic musical-interval continuum: evidence for dual processing. *J. Exp. Psychol. Human*, **9**, 739–752.
- Zatorre, R. & Halpern, A.R. (1979) Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Percept. Psychophys.*, **26**, 384–395.
- Zhang, Y., Kuhl, P.K., Imada, T., Kotani, M. & Tohkura, Y. (2005) Effects of language experience: neural commitment to language-specific auditory patterns. *NeuroImage*, **26**, 703–720.