

University of Memphis

University of Memphis Digital Commons

Electronic Theses and Dissertations

4-18-2016

Stimulus Familiarity and Attentional Effects on the Neural Organization of Auditory Categorical Perception

Breya Symone Walker

Follow this and additional works at: <https://digitalcommons.memphis.edu/etd>

Recommended Citation

Walker, Breya Symone, "Stimulus Familiarity and Attentional Effects on the Neural Organization of Auditory Categorical Perception" (2016). *Electronic Theses and Dissertations*. 1347.
<https://digitalcommons.memphis.edu/etd/1347>

This Thesis is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact khggerty@memphis.edu.

STIMULUS FAMILIARITY AND ATTENTIONAL EFFECTS ON THE NEURAL
ORGANIZATION OF AUDITORY CATEGORICAL PERCEPTION

by

Breya Symone Walker

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Major: Psychology

The University of Memphis

May 2016

Copyright © 2016 Brea Walker
All rights reserved

Acknowledgements

This Master's Thesis would not have been possible without the support and dedication of my thesis committee, Drs. Frank Andrasik, Gavin Bidelman, and Phil Pavlik, Jr. A very special thank you goes to Dr. Gavin Bidelman for teaching me all about event related potentials and allowing me to embark on this journey with his help and guidance.

Abstract

Categorical perception (CP) is the ability to make a comparative judgment on two or more sounds based on stored auditory perceptual information. CP has been shown to be influential on the expression of language proficiency and other cognitive processes such as reading. Prior research has shown that CP is impacted by attention and specific stimulus characteristics, with conflicting results purporting the expression of CP to occur under contradictory conditions. The current investigation examined the expression of neural and behavioral CP under different listening conditions (i.e., passive vs. active tasks) and during engagement with speech and music stimuli. Our results indicate that CP is influenced by attention (active > passive) and is stronger for more familiar stimulus domains (speech > music). Thus, CP does not generalize to other domains in which a listener is not experienced in and it requires a listener to be actively engaged with the auditory stimulus.

Table of Contents

Chapter	Page
1 Introduction	1
2 Methods	7
3 Statistical Analyses	15
4 Results	18
5 Discussion	28
6 Conclusions	33
References	35
Appendices	40

List of Figures

Figure		Page
1.	Spectral characteristics of selected speech stimuli.	10
2.	Spectral characteristics of music stimuli.	12
3.	Standard 10-10 international electrode locations for the EEG	14
4.	Psychometric identification functions for speech and music continua.	19
5.	Comparison of the steepness of listeners psychometric identification functions	20
6.	Behavioral reaction times for speech and music stimuli.	21
7.	Grand average ERP waveforms for prototypical and ambiguous stimuli for speech and music stimuli.	23
8.	Scalp topographic maps of the cortical response to speech and music stimuli in active and passive listening conditions	24
9.	Bar graphs showing the impact of attentional load and stimulus familiarity on cortical ERP activity to speech and music stimuli.	26
10.	Weighted least square regression of psychometric slopes and behavioral responses	28

Stimulus Familiarity and Attentional Effects on the Neural Organization of Auditory Categorical Perception

Categorical perception (CP) is reflective of the ability to make a comparative judgment regarding two or more different stimulus types based on stored perceptual information. CP is an integral aspect of language acquisition (Kuhl, Williams, Lacerda, Stevens, & Lindholm, 1992; Winkler et al., 1999) and an essential part of reading and writing (Werker & Tees, 1984). CP also exists in other domains such as music (Burns & Ward, 1978) and enhancements in CP are conducive to increases in cognitive processes such as second language-learning proficiency (Cooper & Wang, 2012). CP is an innate ability but is also malleable with experience and is evident even during infancy (Kuhl et al., 1992). In the case of audition, CP can be measured with the use of a sound continuum where a single stimulus property (e.g., vowel formant) is parametrically morphed in a continuous manner. Despite continuous variations in the acoustic signal, listeners' perception abruptly changes midway through the continuum. This dichotomous shift in perception is observed in listeners' corresponding psychometric identification functions.. The steepness of the psychometric function is thought to reflect the precision/accuracy in identifying stimulus categories where sharper (i.e., steeper) identification is indicative of more dichotomous categorical perception (Eimas, Siqueland, Jusezyk, & Vigorito, 1971; Pisoni, 1973; Pisoni & Luce, 1987). The plasticity of auditory CP and its influence on perceptual processing warrants further investigation.

CP has been investigated in a variety of domains including speech (Bidelman, Moreno, & Alain, 2013; Bidelman, Weiss, Moreno, & Alain, 2014; Bomba, Choly, & Pang, 2011; Dehaene-Lambertz, 1997; Näätänen & Picton, 1987; Näätänen et al., 1997; Sharma & Dorman, 1999, 2000; Winkler et al, 1999; Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006; Zhang,

Kuhl, Imada, Kotani, & Tohkura, 2005) and music (Burns & Wards, 1978; Cutting & Rosner, 1974; Virtala, Putkinen, Huotilainen, Makkonen, & Tervaniemi, 2012). Prior research has shown that individuals with previous experience to speech (Zhang et al., 2005) or music (Burns & Wards, 1978) sounds through years of exposure and training, respectively, are better able to categorize those sounds compared to individuals who do not have previous exposure. Moreover, acute training enhances categorization of speech and musically complex sounds (Cutting & Rosner, 1974) in college students who did not undergo years of training. Presumably, the salience of CP is not fixed but changes depending on ones' listening experience.

Listeners with a repository of stored auditory information within a specific domain possess the ability to form comparative judgments among between-category stimulus types. Whereas listeners who have less familiarity with the acoustic stimuli of a certain domain (e.g., non-musicians listening to musical intervals) are unable to perceive the subtle distinctions between adjacent stimuli resulting in weaker CP. By assessing CP as it relates to identification between auditory modalities one can infer the degree to which CP exists and what parametric manipulations impact the expression of CP. Further, by assessing CP through identification researchers are able to obtain a conclusive measure of behavioral appraisal. However, identifying differences between stimulus types is not the sole means to obtaining measures of CP.

Past research has utilized event related brain potentials (ERPs) as a way to assess the neural correlates of CP and other auditory perceptual processes (Kappenman & Luck, 2011; Näätänen, 1990; Näätänen & Picton, 1987). ERPs measure time-locked electrical responses of the brain when stimulated by external sensory stimuli (e.g., auditory or visual cues) or perceptual-cognitive events. ERPs are detected by means of time-averaging neural activity recorded via the electroencephalography (EEG). In response to auditory stimuli, the ERPs

consist of a series of obligatory “waves” (components) that develop over the first few hundred milliseconds after stimulus presentation or task execution (Chartrand & Belin, 2006; Habib & Besson, 2009; McMullen & Saffran, 2014; Näätänen et al., 1997; Näätänen, 2001; Pantev & Herholz, 2011; Winkler et al., 1999). In particular, the auditory P1-N1-P2 complex [latency range of ~50-250 ms] has been shown to be sensitive to acoustic properties of sound (e.g., frequency, intensity, etc.) and occurs without the presence of attention (Garinis & Cone-Wesson, 2007; Luck, 2005). Thus, these ERPs are thought to reflect the exogenous processing of sensory information (Luck, 2005).

The P1 response, which is elicited 50 ms post-stimulus, is the first positive going deflection of the auditory cortical ERPs and is thought to reflect the summation of synaptic transmission and arrival of sound in early auditory cortex (Luck, 2005). The N1 response, which is elicited ~100 ms post-stimulus, is the first negative deflection. While the N1 is mainly contingent upon acoustic (exogenous) stimulus properties and signal energy, some studies also suggest it is stronger when participants are attending to stimuli (Chartrand & Belin, 2006), and thus it may reflect early attentional gating of stimulus encoding. Finally, the P2 response [150-275 ms post-stimulus] is sensitive to learning and memory (Habib & Besson, 2009) and therefore, may reflect activation of auditory association regions and sound-to-meaning processing. Importantly, ERPs can be used to assess both pre- and post-perceptual-cognitive processes. Furthermore, these neurophysiological responses make the technique ideally suited to investigate auditory processing and CP from a completely objective manner, which is not possible in traditional behavioral methodology. However, by assessing neural activity in conjunction with behavioral methodology the nebulous brain-behavior relationship can be

elucidated and exogenous/endogenous influences on perceptual processes, such as CP, can be fully examined.

Parametric manipulations have been reported to influence perceptual processes such as CP. In particular, several factors including attention and stimulus familiarity have been shown to impact the neural encoding of auditory stimuli. Many researchers have shown that attention can impact auditory cortical responses (Hillyard, Hink, Schwent, & Picton, 1973; Luck, Heinze, Mangun, & Hillyard, 1990;). For example, previous studies have demonstrated that the N1-P2 response is larger when stimuli are attended to compared to when they are ignored (Hillyard et al., 1973; Luck et al., 1990). Directed attention to stimuli also results in enlargements in the cortical ERPs, particularly the later responses (e.g., N1-P2; Hillyard & Anllo-Vento, 1998). These results suggest that the neural encoding of sound is enhanced during attention. Presumably, attention-related increases in ERP responses result from the recruitment of additional populations of neurons controlling the encoding of stimuli (Luck, 2005).

In addition to attentional effects, stimulus familiarity has been shown to modulate auditory ERPs. Most notable, American and Swedish infants are reported to exhibit language-specific patterns of phonetic perception in response to prototypical native speech sounds (Kuhl et al., 1992) compared to non-native sounds. The early emergence of CP during infancy provides evidence that CP is innate and partially automatic (Kuhl et al., 1992). More recently, the effects of stimulus familiarity on categorical processing have been investigated via cross-language comparisons of CP for speech in native and non-native speakers (Bidelman & Lee, 2015). It was postulated that with long-term experience with lexical tones (as in Mandarin Chinese), native speakers would show enhancements in the neural encoding and CP of pitch compared to non-native speakers who are unfamiliar with the use of lexically-contrastive pitch (i.e., pitch patterns

that change meaning at the word level). Results demonstrated an overall enhancement of the auditory cortical ERPs and stronger categorical coding and perception of linguistic pitch patterns in native compared to non-native speakers. These results support the notion that long-term familiarity with a specific stimulus domain produces experience-dependent changes in auditory brain function and improves the neural encoding and CP for familiar stimuli (Bidelman & Lee, 2015). Collectively, these findings suggest that while CP can be partially automatic (Kuhl et al., 1992) the neural encoding and perception of certain sounds can be enhanced based on listeners' degree of familiarity with certain auditory stimuli (Bidelman & Lee, 2015; Kuhl et al., 1992).

Two recent studies influenced the current investigation by encompassing several, but not all, factors that have been purported to impact CP (e.g., stimulus familiarity and attention). With these factors being assessed singularly, cumulative results regarding the neural representation of CP with regards to the influence of attention and stimulus familiarity cannot be drawn. Chang et al., (2010) assessed CP of speech in the posterior superior temporal gyrus (pSTG) via invasive electrocorticography (ECOG) recordings. Utilizing a morphed speech continua of /ba/ to /da/ to /ga/, neural representations of speech sounds were shown to categorically organize in the pSTG (i.e., just posterior to primary auditory cortex). A clear stimulus boundary was revealed between /da/ and /ga/ in both neural and behavioral responses. Thus, neural responses strongly organized in the early auditory cortices according to their phonetic categories rather than acoustic properties. Two caveats to the experiment of Chang et al. (2010) were (i) the use of a passive listening task and (ii) speech stimuli in order to assess the neural encoding of CP. Consequently, Chang et al., could not conclude what role attention has on the neural encoding of CP nor could they conclude how stimulus familiarity will effect CP. Under Chang's interpretation, it is

suggested that the brain may partially categorize auditory speech stimuli even prior to attentional engagement.

Expanding on the work of the Chang et al. (2010) experiment, Bidelman et al., (2013) investigated the emergence of CP in the auditory system to speech sounds using non-invasive scalp ERPs. Listeners' behavioral psychometric identification functions for an /a-/u/ speech continuum were compared to their "neurometric" functions, derived from their ERPs. Using multivariate analysis (multidimensional scaling and clustering), results of this study showed that listeners' behavioral CP for speech could be accurately predicted based on the dissimilarity of their ERPs evoked by the various tokens along the continuum. Essentially, phonetically similar speech sounds (i.e., within-category tokens) elicited invariant neural activity, whereas phonetically dissimilar tokens (i.e., crossing the categorical boundary) were differentiated in neural responses. However, this study was limited to examining brain responses at a single scalp electrode evoked in an active identification task using speech stimuli. No conclusions could be drawn as to what role attention plays on the neural encoding of CP or whether these findings extend to the auditory categorization of other stimulus domains (e.g., music).

Although researchers have extensively investigated the effects of attention and stimulus familiarity on auditory cortical responses and CP for speech sounds, there is a paucity of work that has investigated these effects of attention and familiarity on the CP for musical stimuli. The current investigation serves as a link, by drawing upon the findings of Chang et al. (2010) and Bidelman et al. (2013), to elucidate whether the process of auditory categorization is contingent upon attention. Further, given that CP occurs in multiple domains, the salience of behavioral and neural categorical processing between speech and music has yet to be examined in a single investigation and with the same listeners.

Our current study aimed to more directly investigate the effects of (i) attention and (ii) stimulus familiarity on the neural categorization of auditory information and reconcile equivocal findings in the CP literature. Comparisons between active and passive tasks allowed us to examine whether or not categorical processing in the auditory system requires active engagement with stimuli (e.g., Bidelman et al., 2013) or instead, reflects automatic, pre-attentive grouping mechanisms (e.g., Chang et al., 2010). Comparisons between music and speech stimuli allowed us to clarify the salience of CP for different classes of auditory stimuli and whether CP is domain specific to speech or if the process is generalized more broadly to other sounds. We hypothesized that cortical neural activity would be modulated by both the stimulus domain (i.e., music or speech) and attentional state (i.e., passive listening or active listening). More specifically, we anticipated finding that auditory stimuli more familiar to our listeners (e.g., speech) would produce clearer categorical-like brain responses (e.g., Bidelman & Lee, 2015) compared to sounds for which listeners have not yet developed categorical labels (i.e., musical intervals; Pantev et al., 1998). Additionally, we expected to find an effect of attentional load, such that categorical neural encoding would only emerge when listeners actively attended (and were classifying) auditory stimuli (cf. Bidelman et al., 2013 and Chang et al., 2010).

Methods

Procedural Overview

Following written informed consent, participants completed a questionnaire to obtain information regarding general demographics (i.e., age, gender, past CNS injury/disease), language experience (i.e., native language and second languages spoken), musical experience (i.e., musical instruments played, years per instrument, hours of practice/formal classes), and handedness (i.e., Edinburg Handedness Inventory; Oldfield, 1971).

Participants then completed the active and passive conditions of the experiment inside an electroacoustically shielded booth while their EEGs were recorded. During the active condition, participants identified either music or speech tokens (separate blocks) by responding via a button press on a keyboard (task described below). During the passive condition, participants listened passively to identical music or speech tokens but watched a closed caption movie to create a calm yet wakeful state (Bidelman et al., 2014). Prior to each run, instructions were given contingent upon the condition presented. The active condition emphasized speed of identification (e.g., “*Respond as fast as you can and refrain from moving.*”), while the passive condition emphasized inattentiveness (e.g., “*Ignore the sounds you hear and refrain from moving.*”). Total testing time took approximately 2.5 hours and participants were compensated \$20/hr for their participation. The study was approved by the Institutional Review Board at the University of Memphis (IRB #2370, Appendix C).

Participants

Ten young adults (1 male, 9 females; age: $M = 22.5$, $SD = 2.8$ years) were recruited via fliers from the University of Memphis to participate in the experiment. All participants exhibited normal hearing sensitivity (i.e., < 25 dB HL) as determined by an audiometric screening at octave frequencies between 500 and 2000 Hz. All were native speakers of American English and reported no previous history of neurological diseases/injuries. Each listener was strongly right-handed (78% laterality index; Oldfield, 1971) and had minimal formal musical training ($M = 0.6$, $SD = 0.8$ years). Age-related hearing loss (Alain, Roye, & Salloum, 2014; Bidelman, Villafuerte, Moreno, & Alain 2014), tone-language experience (Bidelman, Gandour, & Krishnan, 2011), and musical training (Bidelman et al., 2011; 2014; Zendel & Alain, 2013) are known to

modulate the auditory cortical ERPs and CP. To avoid these confounds in the present study, individuals were excluded from participation if they reported a 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.

Speech Stimuli

We used a synthetic five-step vowel continuum from /u/ to /a/ to assess CP for speech (Bidelman et al., 2013, 2014). Each token of the continuum was separated by equidistant steps acoustically based on first formant frequency, yet was perceived categorically. Each token was 100 ms, including 10 ms of rise/fall time to reduce spectral splatter in the stimuli (Bidelman et al., 2013, 2014). Tokens contained 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., 2013, 2014). Spectral characteristics of the speech stimuli are shown in Figure 1.

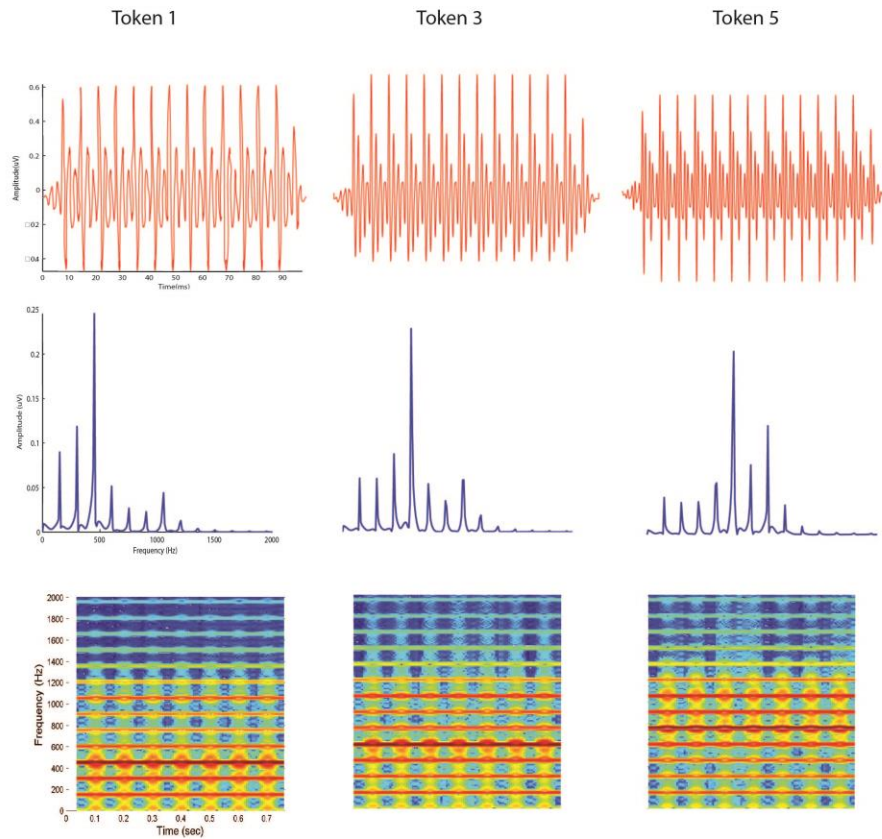


Figure 1. Spectral characteristics of selected speech stimuli. Shown here are time series, frequency spectra (FFTs), and spectrograms for speech tokens 1, 3 and 5, reflecting the ends (Tk 1, 5) and midpoint (Tk 3) of the vowel continuum.

Music Stimuli

As with speech, we synthesized a continuum of musical intervals to assess the CP for musical sounds. 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 harmonics with equal amplitudes added in cosine phase. For each musical token, the lower of the two pitches was fixed with a fundamental frequency (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 and are hence the most familiar to both musicians and non-musicians (Brattico et al., 2008). Moreover, they connote the typical valence of “sadness” (m3) and “happiness” (M3) and are thus easily described to participants unfamiliar with music-theoretic labels. Acoustic characteristics for the music stimuli are shown in Figure 2.

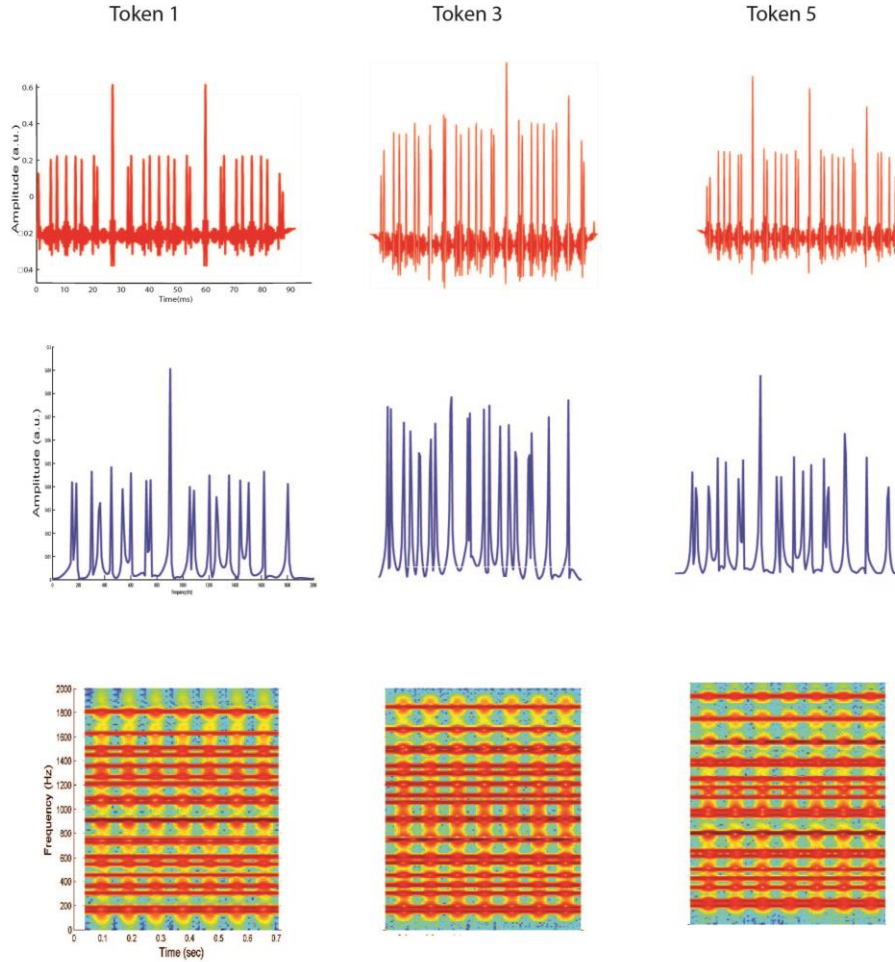


Figure 2. Spectral characteristics of music stimuli. Shown here are time series, frequency spectra (FFTs), and spectrograms reflecting the ends (Tk 1,5) and midpoint (Tk 3) along the music continuum.

Stimulus Presentation and Task Conditions

A total of eight active/passive conditions were collected for each participant, with the active conditions requiring identification of stimuli via button presses. The passive condition did not require identification of stimuli; participants watched a silent, self-selected movie during stimulus presentation and were instructed to ignore the sounds they heard. The condition blocks were presented in alternating order of active and passive blocks. There were a total of two runs of the music continua (each 100 trials * 5 tokens) and two runs of the speech continua (each 100

trials * 5 tokens) during the active condition. Similarly, the four passive conditions consisted of two runs of each of the speech and music stimulus sets where listeners heard 500 trials per continuum. In total, listeners heard 200 trials of each individual speech/music token presented in either an active or passive listening condition.

Stimulus presentation was controlled by MATLAB® 2013 (The MathWorks, Inc.) routed through a TDT RP2 interface and HB7 headphone buffer (Tucker-Davis Technologies). Stimuli were delivered binaurally at an intensity of 83 dB SPL through insert earphones (ER-2, Etymotic Research). Stimulus intensity was calibrated using a Larson-Davis SPL meter (Model 824, Provo, Utah) measured in a 2-cc, artificial ear coupler (Model AEC1001). Left and right ear channels were calibrated separately. In the active condition, stimuli were presented with an inter-stimulus interval (ISI) that was randomly jittered between 400 and 600 ms (uniform distribution) following the collection of the participant's behavioral response. The auditory cortical ERPs experience a rapid attenuation for faster presentation rates (Bidelman, 2015; Picton, Woods, Baribaeu-Braun, & Healy, 1977). Hence, for passive condition, 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.

ERP Acquisition

Neuroelectric activity was recorded from 64 sintered Ag/AgCl electrodes at standard 10-10 locations around the scalp (Oostenveld & Praamstra, 2001; Figure 3) using procedures described in previous reports (e.g., 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 eye movements (i.e., blinks and saccades). 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 (CAR). Contact impedances were maintained < 10 k Ω during data collection.

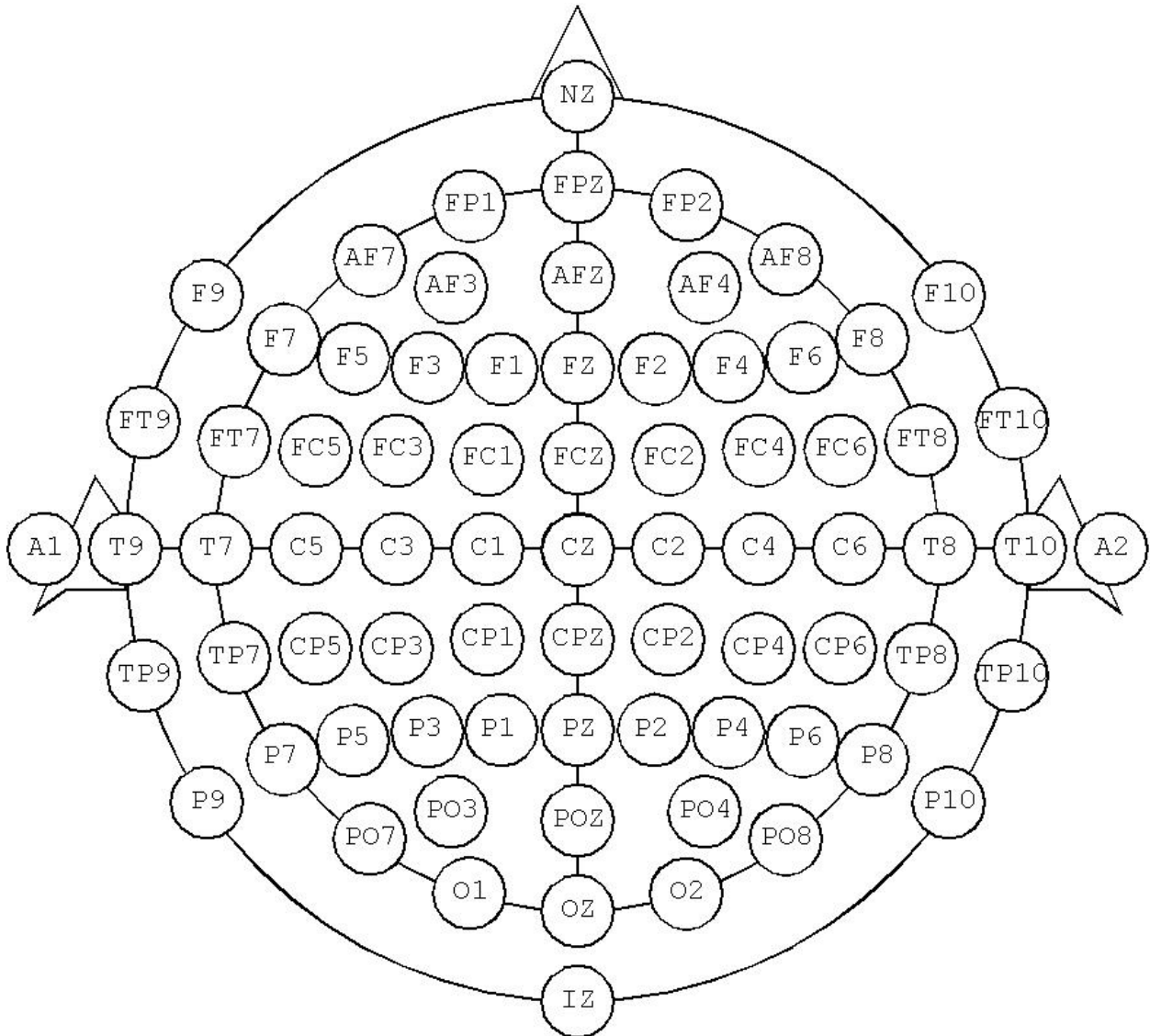


Figure 3. Standard 10-10 international electrode locations for the EEG

(<http://frontalcortex.com/images/eeg/1010labels.jpg>).

ERP Data Preprocessing

Preprocessing was performed in Curry 7 (Compumedics Neuroscan) and custom coded MATLAB scripts. Ocular artifacts were corrected in the continuous EEG using principal

component analysis (PCA). Statistically, PCA increases the signal-to-noise ratio within EEG data that can be reduced by ocular and myogenic artifacts (e.g., blinks). This results in a set of averaged principal components (PCs) describing variance in the EEG time series. Ocular artifacts typically load into the first few PCs given their large ($> 50 \mu\text{V}$) amplitude compared to the ERPs of interest ($< 1\text{-}2 \mu\text{V}$). These artifactual PCs were then subtracted from the continuous EEG data, nullifying their influence on the recordings. Cleaned EEGs were then digitally filtered (0.1-30 Hz; zero-phase filters), epoched (-200-800 ms, where $t = 0$ was the onset of the stimulus), base-line corrected to the pre-stimulus period, and averaged in the time domain to obtain ERPs for each stimulus. This time-locked averaging resulted in 20 ERP waveforms per participant (5 tokens * 2 listening conditions * 2 stimulus domains).

Statistical Analyses

Behavioral Data Preparation

For the active condition, participants' behavioral responses to speech/music tokens were logged and both %-identification of each token and reaction times (RTs) were recorded (only in the active listening condition). Faster RTs and steeper psychometric identification curves are indicative of faster labeling speeds and more precise (dichotomous) identification of the stimulus continuum (i.e., more salient CP; Bidelman et al., 2014). Psychometric identification functions were constructed for each stimulus domain and participant by computing the proportion of trials listeners identified as one category or the other. Given that there was only a binary option for identification, the analysis resulted in two complementary identification functions (for each domain).

Psychometric identification scores were fit with a two-parameter sigmoid function. We utilized a logistics function: $P = 1 / [1 + e^{-\beta_1(x-\beta_0)}]$ where P is the proportion of trials

identified as a given stimulus token, x the step number along the stimulus continuum, and β_0 and β_1 the location and slope of the logistic fit, respectively (Bidelman et al., 2014). This allowed us to measure the location and steepness (i.e., slope) of the CP boundary, where listeners' perception abruptly shifted categories. Fits were estimated using an iterative, nonlinear least-squares regression procedure as implemented in MATLAB's 'nlinfit' function. An independent samples t -test was utilized to examine the difference between β_1 values between domains. Larger β_1 values reflect steeper psychometric functions and hence, would indicate stronger categorical perception in one stimulus domain versus the other.

Behavioral stimulus-labeling speeds (i.e., RTs) were computed as the listener's median response latency across all trials for a given condition. One participant's RT data was not recorded during the active music task due to hardware malfunction. Consequently, nine participants' RT data was utilized during analysis of music RTs. To be consistent with our previous reports using similar CP tasks (Bidelman et al., 2013, 2014), RTs outside of 250-2500 ms were classified as outliers (i.e., improbably fast, lapses of attention) and excluded from further analyses. Behavioral RTs were pooled for tokens at the ends (Tk 1, 5) and midway (TK 3) through the continua in order to increase statistical power. Tk 1,5 represents true dichotomous categories, while Tk 3 represents an ambiguous category. From here on, Tk 1,5 will be referred to as prototypical tokens and Tk 3 will be referred to as ambiguous tokens.

ERP Response Preparation

For the purpose of data reduction and to minimize potential bias in electrode selection for data analysis, we collapsed a subset of the 64-channel sensor data into a single region of interest (ROI) electrode cluster encompassing several scalp electrodes located at the front of the scalp. The resulting ROI consisted of the average of six fronto-central electrodes (F1, Fz, F2, FC1,

FCz, FC2). This ROI was utilized for data analysis given that the auditory cortical responses are best represented at fronto-central locations on the scalp (Bidelman et al., 2013, 2014; Bidelman & Lee, 2015).

P1-N1-P2 wave latency and amplitudes were identified from each cortical ERP waveform as the peak positivity/negativity at ~50 ms, 80-100 ms, and 150-275 ms post stimulus presentation, respectively. However, to provide a single, parsimonious measure of overall cortical responsiveness to our stimuli, we computed N1-P2 magnitudes, derived as the difference between the individual N1 and P2 wave amplitudes. Previous research has shown that N1 and P2 ERPs are sensitive to speech perception tasks (Alain et al., 2007, 2010; Bidelman et al., 2013; Wood et al., 1971) and are prone to neuroplastic effects associated with speech and music sound training (Alain et al., 2007; Seppanen et al., 2012; Shanin et al., 2003; Tremblay, Kraus, Ponton, & Otis, 2001). Thus, we restricted the majority of our ERP analyses to N1-P2 magnitudes (measured in μV).

Unless otherwise noted, behavioral and neural response measures were analyzed separately by stimulus domain (speech, music) using mixed model ANOVAs (subjects= random factor) with fixed effects consisting of listening condition (active vs. passive) and token type (prototypical tokens vs. ambiguous tokens) (PROC GLIMMIX, SAS® 9.4; SAS Institute, Inc.). An *a priori* significance level was set at $\alpha = 0.05$. Multiple comparisons were corrected via Holm-Tukey adjustments.

Brain-Behavior Relations

To examine the degree to which neural representations for speech and music stimuli predict the behavioral categorization, we performed weighted least square regression between listeners' neural responses and their β_1 values (i.e., psychometric slope). Psychometric slopes

(β_1) reflect the degree to which listeners distinguish prototypical from ambiguous speech categories. 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 N1-P2 magnitude), computed as the *difference* between the average N1-P2 magnitudes evoked by the Tk1/5 tokens and ambiguous case (Tk3). We then regressed Δ N1-P2 against behavioral β_1 responses (i.e., steepness of the psychometric function). This regression analysis reveals the degree to which changes in neural activity between prototypical vs. ambiguous sounds reflect their successful behavioral categorization.

Results

Behavioral Identification

Behavioral identification functions are shown in Figure 4. Comparisons between speech and music psychometric functions show dichotomous CP for speech stimuli along the continuum, with an abrupt shift in the perceptual category occurring midway through the continuum. In contrast, the music continuum shows 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. 5), which revealed steeper identification (larger β_1) when identifying speech compared to music [$t(9) = 26.868, p < 0.001$]. These findings suggest that the salience of CP is influenced by an individual's familiarity with the stimulus set; participants in our sample had minimal formal musical training and were thus, less familiar with the categorical labels of musical intervals. More familiarity with speech stimuli resulted in more precise CP between categories compared to the less familiar stimulus domain of music.

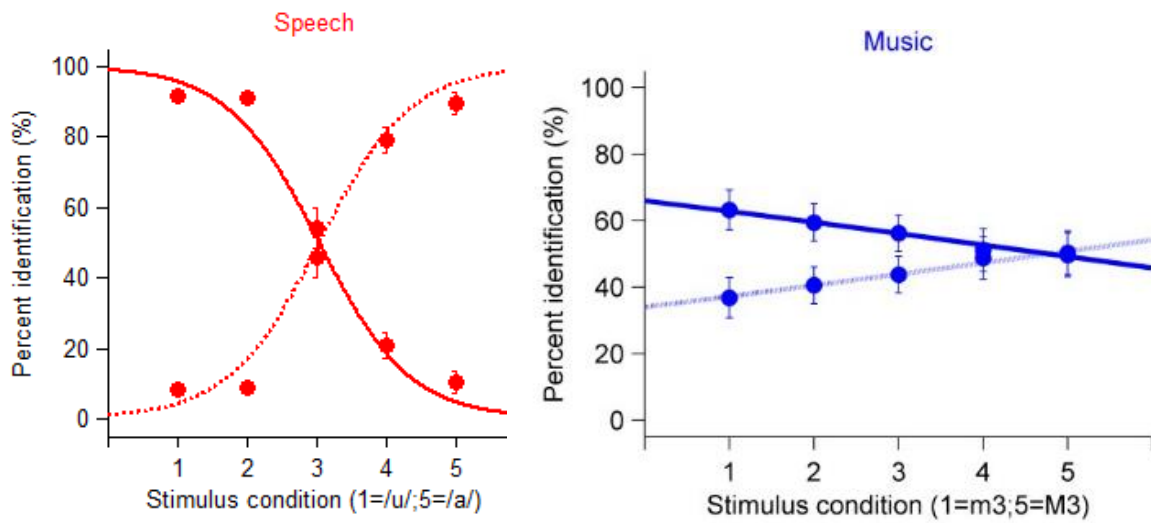


Figure 4. Stimulus familiarity enhances CP. Psychometric identification functions for speech (*left*) and music (*right*) continua. Data points represent averaged percent identification for each stimulus token along the continuum with error bars displaying the standard error for each token. Solid lines represent the psychometric function from /u/ to /a/ for speech and m3 to M3 for music. Dotted line shows the inverse psychometric function for each stimulus domain from /a/ to /u/ and M3 to m3.

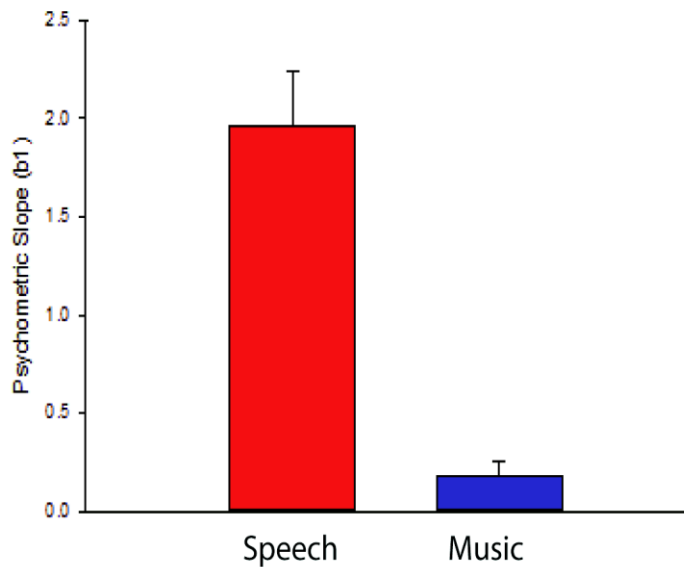


Figure 5. Comparison of the steepness (i.e., slope= β_1) of listeners psychometric identification functions when classifying speech and music stimuli (see Fig. 4). Error bars represent the standard error of the mean.

Behavioral RTs for classifying stimuli from the speech and music domain are shown in Figure 6. Qualitative comparisons between speech and music RTs suggest that speech stimuli were labeled faster than music stimuli. However, we aimed to determine if RTs were faster for prototypical compared to ambiguous tokens, which is characteristic of salient CP (Pisoni & Tash, 1974; Bidelman et al., 2013, 2014). To this end, we contrasted RTs for Tk1/5 and Tk3 for each domain. A paired samples *t*-test (two-tailed) revealed a significant difference between RTs in prototypical speech ($M = 427.1$, $SD = 96.5$) vs. ambiguous speech ($M = 490.1$, $SD = 101.1$) categories [$t(9) = 3.49$, $p = 0.007$]. There was also a significant difference between RTs in prototypical musical tones ($M = 478.1$, $SD = 142.7$) vs. ambiguous musical tone ($M = 456.2$, $SD = 128.6$) categories [$t(8) = -2.96$, $p = 0.018$]. Overall, prototypical stimuli (endpoint tokens) resulted in faster RTs compared to ambiguous (midpoint) stimuli for both speech and music.

Furthermore, while this RT effect appeared stronger for the speech continuum, it was the only domain to show true dichotomous identification as music stimuli were perceived in a continuous manner (see Fig. 4).

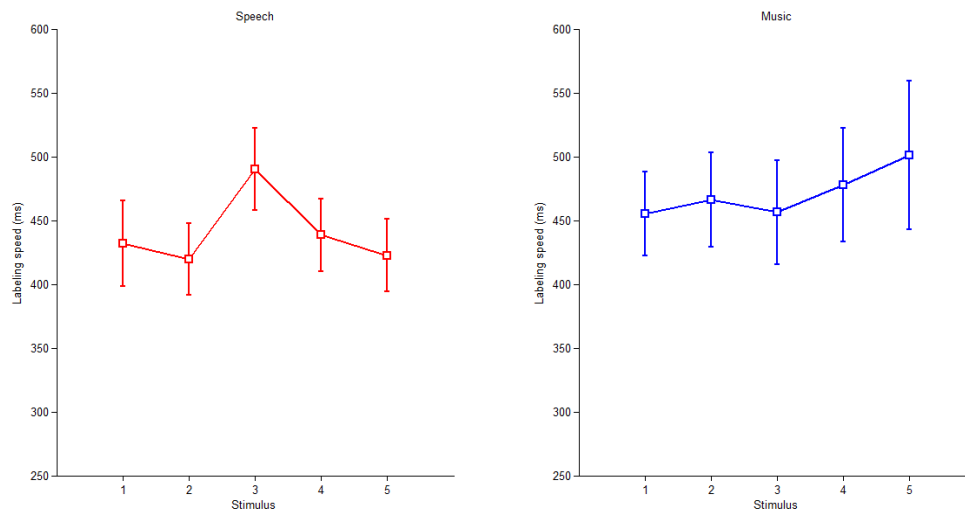


Figure 6. Behavioral RTs for speech (*left*) and music (*right*) stimuli. For speech, RTs to prototypical stimuli are faster than music RTs to prototypical stimuli, indicative of salient CP. In contrast, no difference in RTs were found for classifying musical stimuli, indicative of weak/absent CP (i.e., continuous perception). Error bars display the standard error of the mean.

Cortical ERPs

Grand average ERPs waveforms by stimulus domain and category are shown in Figure 7. Scalp topographic maps are shown in Figure 8. Grand average ERPs show overall neural activity at the clustered fronto-central electrode locations for prototypical and ambiguous tokens for each stimulus domain. Topographic scalp distributions show overall neural activation during active and passive listening for prototypical vs. ambiguous tokens.

An ANOVA conducted on individual latencies of the ERP waves (i.e., P1-N1-P2) revealed no differences in response latency with the experimental manipulations (Appendix A). In stark contrast, significant modulations in ERP amplitudes were observed with attention and

category type (i.e., prototypical vs. ambiguous tokens). For speech, N1 wave amplitudes did not differ with attention [$F(1, 9) = 1.75, p = 0.22$] or category type [$F(1, 9) = 1.72, p = 0.22$], but there was a significant interaction [$F(1, 9) = 5.38, p = 0.045$]. Follow-up contrasts revealed that the neural encoding of ambiguous tokens was impacted by attention [$t(9) = 2.29, p = 0.048$]. There was a significant difference of neural encoding of prototypical and ambiguous speech stimuli during active listening [$t(9) = -2.57, p = 0.03$].

For music, N1 wave amplitudes did not differ by stimulus category [$F(1, 9) = 0.00, p = 0.99$], listening condition [$F(1, 9) = 2.50, p = 0.15$], or their interaction [$F(1, 9) = 0.02, p = 0.89$]. Similarly, individual P2 wave amplitudes did not differ for either speech or musical stimuli with changes in listening condition [speech: $F(1, 9) = 1.97, p = 0.19$; music: $F(1, 9) = 0.00, p = 0.95$], stimulus category [speech: $F(1, 9) = 0.07, p = 0.79$; music: $F(1, 9) = 2.31, p = 0.16$], or their interaction [speech: $F(1, 9) = 0.74, p = 0.41$; music: $F(1, 9) = 1.46, p = 0.25$].

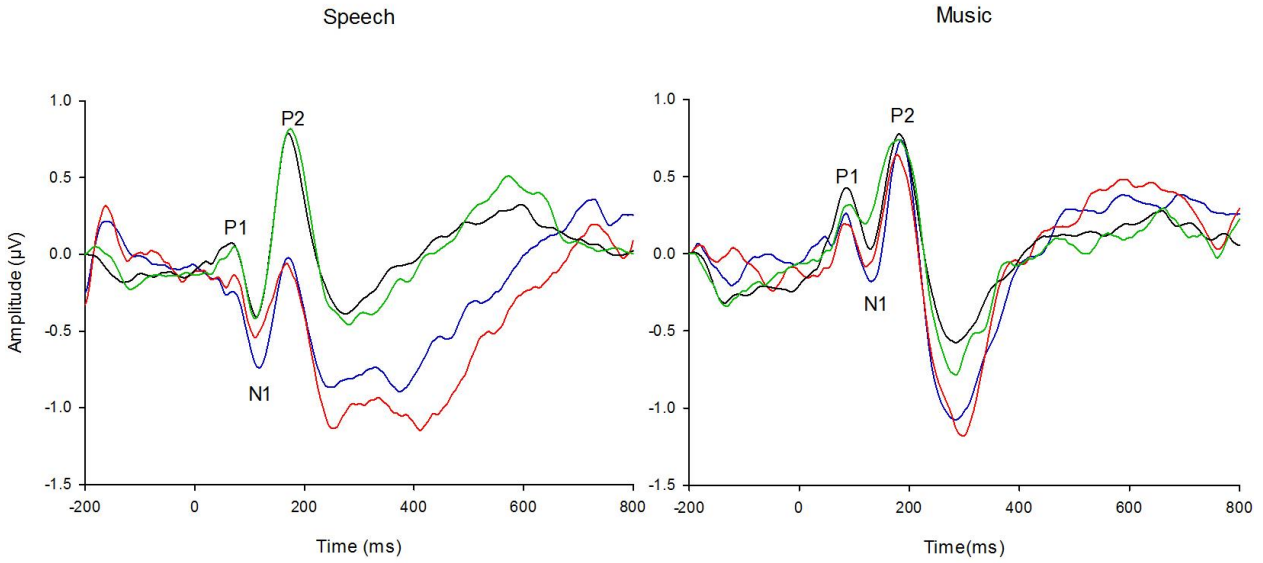


Figure 7. Grand average ERP waveforms for prototypical and ambiguous stimuli for speech (*left*) and music (*right*) stimuli. Traces represent ERPs to prototypical stimuli during active (blue) and passive (black) listening and ambiguous stimuli during active (red) and passive (green) listening. Onset of stimulus presentation is represented by time=0. Prominent ERP deflections (labeled), within the traces prior to ~400ms represent the P1-N1-P2 waves that were analyzed.

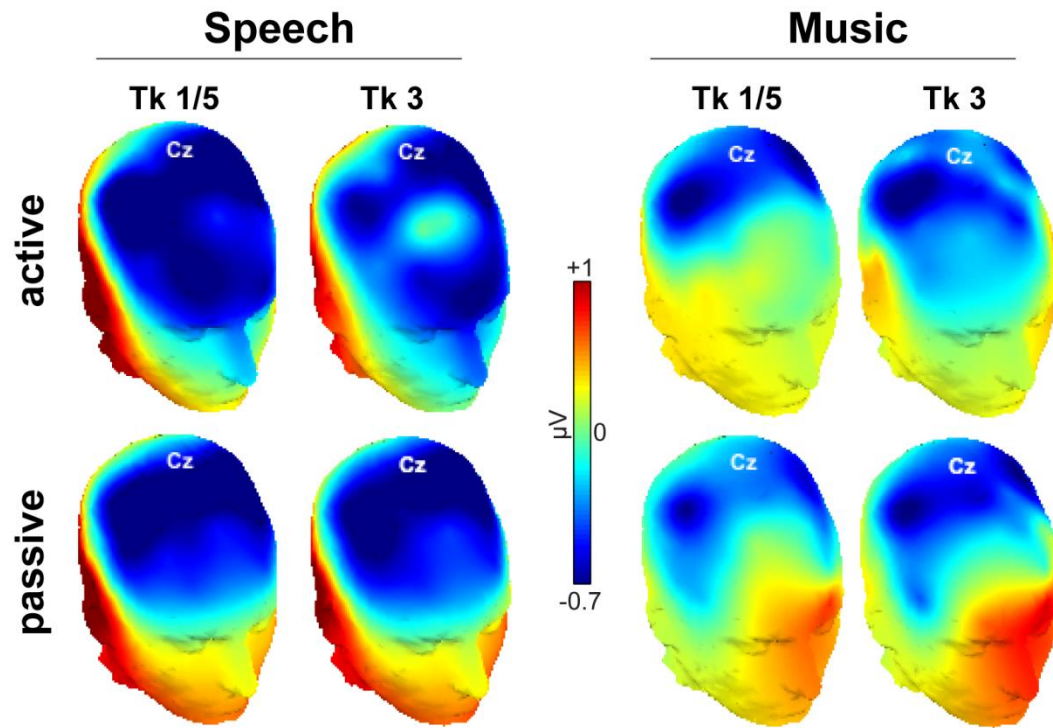


Figure 8. Scalp topographic (isopotential) maps of the cortical response to speech and music stimuli in active and passive listening conditions. Hot and cool colors denote positive and negative voltage, respectively.

In the present study, active and passive stimulus blocks were presented serially to limit variability in subject state across blocks. However, one concern with this sequencing is that neural responses collected later in the session could show habituation effects due to stimulus presentation order (e.g., active trials > passive trials). The P1 component has been reported to be sensitive to stimulus properties including presentation rate and is thought to reflect synaptic events coding the arrival of information within the auditory cortex (Nash, Sharma, Martin, & Biever, 2008). Thus, the P1 allowed us to investigate whether block order had any effect on the overall activation level (note also that our stimuli were identical regardless of task). In theory, P1 responses during later experimental blocks (i.e., the passive task) could be weaker than active listening blocks.

To rule out neural fatigue effects due to block order an ANOVA was conducted on the P1 amplitudes with factors of listening condition and category type. We found no significant differences of the P1 response with listening condition ($F(1,9) = 1.79, p = 0.21$), stimulus category ($F(1,9) = 1.11, p = 0.32$), or their interaction ($F(1,9) = 2.95, p = 0.12$). This suggests that fatigue or order effects did not play a role in the observed changes in N1 and P2. This is further supported by the fact that ERPs to speech and musical stimuli showed differential attentional effects (*speech*: attentional modulations; *music*: attentional invariance; discussed below). It would be difficult to account for these effects based on order or fatigue effects given that speech and music stimuli were randomly interleaved throughout the experiment.

N1-P2 magnitudes, representing our primary measure of overall auditory neural activity to speech/music stimuli, are shown in Figure 9. Generally speaking, we observed that N1-P2 responses varied according to the stimulus category (prototypical vs. ambiguous tokens) and attentional state for speech but not musical stimuli. For speech, analysis of the N1-P2 amplitudes revealed a significant interaction between listening condition and stimulus category [$F(1, 9) = 5.97, p = 0.037$]. Multiple comparisons revealed an attentional effect for the ambiguous speech (Tk 3), whereby ERPs were stronger in the passive compared to active task [$t(9) = -3.08, p = 0.013$]. In contrast, prototypical vowels (Tk1/5) evoked similar N1-P2 amplitudes regardless of attentional state [$t(9) = -0.52, p = 0.613$]. Comparisons by attentional state revealed that N1-P2 amplitudes were (marginally) larger in response to Tk1/5 than Tk3 [$t(9) = 2.15, p = 0.059$]. This indicates that brain activity differentiated prototypical from ambiguous speech sounds producing strong and weak categorical percepts, respectively. In contrast, we found no difference between Tk1/5 and Tk3 N1-P2 amplitudes during passive listening [$t(9) = -1.12, p = 0.2934$]. These

findings indicate that brain activity did not differentiate prototypical from ambiguous speech sounds during passive listening.

In contrast to the attentional and stimulus modulations observed in the ERPs for speech, in the music domain, we found no significant effects of listening condition [$F(1, 9) = 0.01, p = 0.93$], stimulus category [$F(1, 9) = 0.00, p = 0.99$], nor interactions [$F(1, 9) = 0.40, p = 0.54$].

Collectively, these findings indicate that neural correlates of CP are most evident during active listening conditions and for familiar stimuli (i.e., speech > music).

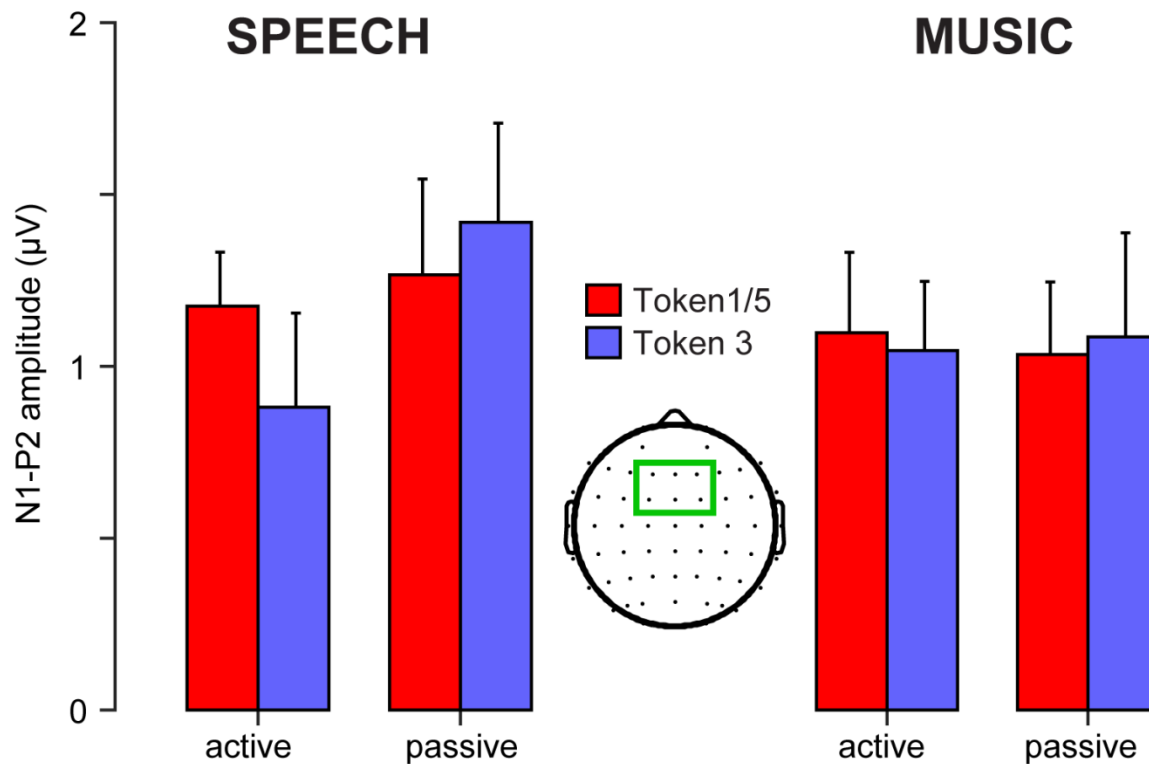


Figure 9. Bar graphs show the impact of attentional load and stimulus familiarity on cortical ERP activity to speech and music stimuli. The inset shows the fronto-central electrode ROI utilized to derive the N1-P2 magnitude. Prototypical stimuli result in stronger categorical processing compared to ambiguous responses primarily in the active speech listening conditions. No attentional or category differences were revealed for music stimuli. Error bars represent standard error of the mean.

Brain-Behavior Relations

Weighted least squares regression was performed in order to examine the correspondence between behavioral and neural responses (Figure 10). Psychometric slopes (β_1) were regressed against cortical $\Delta N1-P2$ magnitudes [i.e., (Tk1/5) - Tk3]. This allowed us to assess the degree to which changes in neural activity between prototypical vs. ambiguous sounds reflect their successful behavioral categorization. Active and passive condition ERPs were considered in the regression in order to see the influence attention had on behavioral identification. Data from both music and speech conditions were pooled to achieve adequate sample size for regression ($N > 10$ observations; Babyak, 2004). The regression analysis revealed that the steepness of listeners' psychometric functions was significantly predicted by larger $\Delta N1-P2$ responses elicited during active listening ($R^2 = 0.38$, $F(1, 19) = 10.97$, $p = 0.004$). This was not evident during the passive listening condition ($R^2 = 0.05$, $F(1,19) = .855$, $p = 0.367$). These findings reveal that the degree to which brain activity distinguishes prototypical from ambiguous tokens predicts the salience of CP listeners experience behaviorally. Moreover, this relationship only exists during active listening.

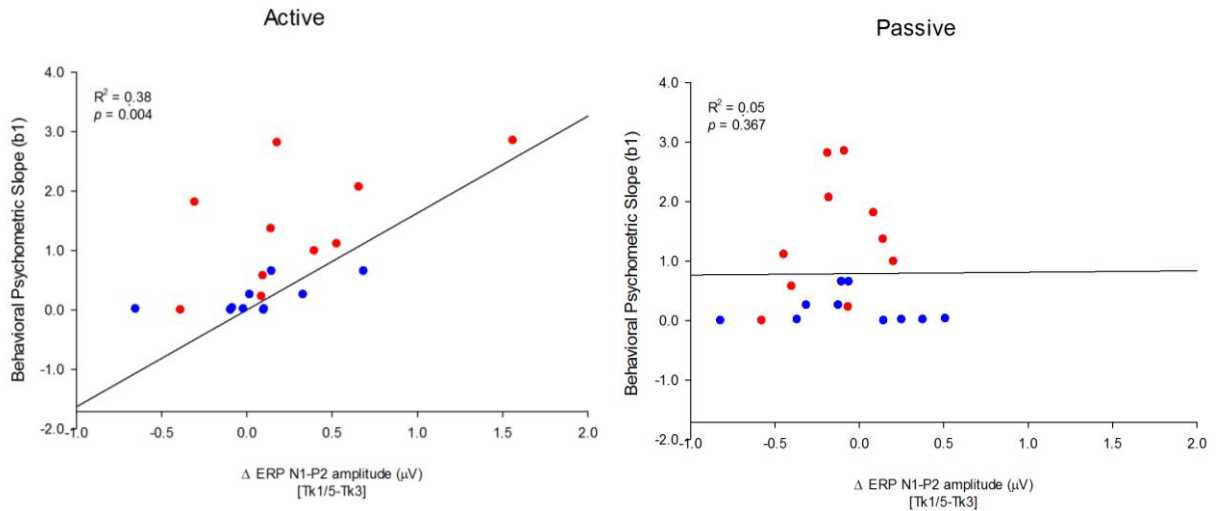


Figure 10. Weighted least square regression revealed that psychometric slopes (β_1) corresponding to behavioral responses are significantly predicted by Δ N1-P2 magnitudes (Tk1/5-Tk3) during active (left) but not passive listening (right). Increases in overall Δ N1-P2 magnitude correspond to increases in β_1 values during active listening. Data points represent individual participants' neural vs. behavioral responses to speech (red) and music (blue).

Discussion

Categorical perception (CP) reflects the ability to make a comparative judgment regarding two or more different stimulus types based on stored perceptual information. CP is an integral aspect of many cognitive processes such as language acquisition (Kuhl et al., 1992; Winkler et al., 1999) and reading and writing (Werker & Tees, 1984). Prior research has shown that CP is expressed for speech (Bidelman et al., 2013, 2014; Bombaet al., 2011; Dehaene-Lambertz, 1997; Näätänen & Picton, 1987; Näätänen et al., 1997; Sharma & Dorman, 1999, 2000; Winkler et al., 1999; Ylinen, et al., 2006; Zhang, et al., 2005) as well as music (Burns & Wards, 1978; Cutting & Rosner, 1974; Virtal, et al., 2012) at both the behavioral and neural level. Furthermore, attentional and stimulus familiarity effects have been reported to impact neural representations of CP (Hillyard, 1973; Hillyard & Anllo-Vento, 1998; Luck, et al., 1990; Näätänen, 2001, Näätänen, et al., 2007; Näätänen & Picton, 1987; Sussman, 2007; Sussman, et

al., 2003; Winkler, 2007). Neural activity has been shown to be larger when stimuli are attended to than when they are ignored (Hillyard et al., 1973; Luck et al., 1990). Our current investigation was inspired by two experiments that assessed the expression of CP to speech sounds during differential listening conditions. Chang et al. (2010) reported neural representations of speech sounds (assessed via near-field intracranial recordings) to be categorically organized during passive listening, whereas Bidelman et al. (2013) reported neural representations of CP (assessed via the scalp ERPs) to be present during active listening.

Due to the conflicting evidence of attentional influences on the neural correlates of CP, the current study attempted to serve as a link to elucidate the influences of attention on the expression of CP. Moreover, while prior research has reported CP for speech (Zhang et al., 2005; Bidelman & Lee, 2015) and music (Virtal et al., 2012) independently, no study has directly compared the degree to which the saliency of CP is impacted by stimulus domain *within the same listeners*. Our results reveal that the expression of CP is dependent upon attentional load, suggesting that CP only occurs during active listening (Figure 9). That is, categorical neural encoding does not seem to occur automatically, during passive listening (cf. Chang et al., 2010). Further, we reveal that the saliency of CP is influenced by prior experience with the stimulus domain (Figures 4 & 9). We found that CP was more prominent while listeners engaged with speech rather than music sounds, consistent with their long-term experience with the former and lack of formal labels for musical sounds (all listeners were nonmusicians). Moreover, prototypical sounds elicited stronger CP in both neural (Figure 9) and behavioral (Figures 4 & 6) responses compared to ambiguous sounds further corroborating the claim that CP is dependent upon stimulus familiarity. Finally, regression analyses revealed that behavioral identification is

associated with neural activity during active (but not passive) engagement with stimuli (Figure 10).

Is the Expression of CP Contingent upon Attentional State?

Our results demonstrate that the neural correlates of CP require active engagement of the listener; categorical neural encoding is not elicited under strictly passive listening. With a number of studies reporting that neural correlates of perceptual processes are impacted by attentional load (Hillyard et al., 1973; Luck, et al., 1990) and with conflicting results in the CP-ERP literature (Chang et al. 2010; Bidelman et al., 2013), our current investigation helps to clarify the relationship between attention and auditory categorical brain processing. Chang et al. (2010) reported CP to be present during passive listening to speech while Bidelman et al. (2013) reported CP to be present during active listening. Our current investigation specifies our prior investigation (Bidelman et al., 2013) by showing that CP is expressed during active listening only. A plausible explanation for Chang et al's reporting of CP during passive listening could be attributed to covert listening. In that study, attention was not directly manipulated nor controlled. With the lack of a distraction task during their passive listening paradigm, it is possible that their results could partially be influenced by participants inadvertently shifting attention while listening to the speech sounds. In order to control for this confound in the present study, we employed a distraction during our passive listening task. This allowed us to measure a more veridical ERP response during passive listening by decreasing the chances of covert listening effects. By doing so, we failed to find neural correlates of CP during passive listening and infer that categorical speech processing strictly requires goal-directed attention.

Is the Saliency of CP Dependent upon Stimulus Familiarity?

Our results demonstrate that CP depends on a listener's prior experience within a stimulus domain and it does not generalize broadly to other sounds. It is suggested that strong neural and behavioral categorical processing requires attention and familiarity to stimuli. Our sample consisted of individuals with no prior musical training. Consequently, their lack of experience with the formal labels of pitch intervals influenced the saliency of their categorization for the musical stimuli. Prior research has shown that individuals with previous experience in a specific auditory domain show stronger categorization of those stimuli compared to those without prior experience (Burns & Wards, 1978; Cutting & Rosner, 1974; Zhang et al., 2005). Our results here are thus consistent with findings reported in infant studies (Kuhl et al., 1992; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; McClelland, Fiez & McCandliss, 2002) and language studies (Bomba et al., 2011; Kirmse et al., 2008; Werker & Tees, 1984; Zhang et al., 2005). For example, infants have been shown to hone in on language properties utilized more frequently within their native language within the first six months of life (Kuhl et al., 1992). With this pruning of unused language properties, infants develop categorical-like responses to language properties that are readily used within their native language compared to those that are not (Kuhl et al., 1992). Neural correlates of CP have also been shown to be smaller or absent in non-native speakers who lack familiarity with aspects of their second language (Bidelman & Lee, 2015; Bomba et al., 2011; Zhang et al., 2005). Mandarin speakers show enhancements in neural activity to pitch sounds compared to non-Mandarin speakers due to the significance of lexically contrastive pitch within the Mandarin language (Bidelman & Lee, 2015). Furthermore, musical training has been shown to impact behavioral identification of musical intervals, such that increases in accuracy and identification are seen in musicians and not

in non-musicians (Burns & Wards, 1978; Cutting & Rosner, 1974). Although the aforementioned studies utilized differing 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. Thus, our current study corroborates the claim that the saliency of CP relies on one's familiarity with a prescribed stimulus domain.

Do Neural Correlates of CP Predict Behavioral Identification?

We examined the relationship between neural correlates of CP and behavioral identification accuracy to elucidate the relationship between neural activity and behavioral sound categorization. Our results demonstrate that neural representations of CP as indexed by the early auditory ERPs predict listeners' behavioral CP accuracy (Figure 10). By regressing ERP $\Delta N1-P2$ magnitudes with the slopes of listener's psychometric identification functions, we provide evidence that neural representations indicative of changes in brain activity between prototypical and ambiguous sounds correspond with the degree of perceptual salience between sound categories (i.e., steeper psychometric functions); increases in neural response magnitudes were linked to increases in behavioral CP. Further, we provide evidence that this relationship between changes in brain activity and behavioral appraisal is only present during *active* listening. We did not find any correspondence between behavioral auditory categorization and ERP coding for the passive listening condition. Collectively, these findings suggest an intimate link between the neural processing of sound and behavioral categorization skills. However, this brain-behavior connection requires one to be actively engaged during categorization.

Limitations and Future Research

While we have provided new evidence for stimulus familiarity and attentional load effects on the expression and neural correlates of CP, no study goes without limitations. We assessed individuals who were native speakers of American English and individuals were largely naïve to musical intervals. Thus, we cannot evaluate how different language experiences or musicianship might impact categorical processing during active and passive listening. However, prior research would suggest that groups who are familiar with acoustic properties of the stimuli in question (i.e., trained musicians, native speakers) would display advantages over groups who are unfamiliar, both behaviorally and neurally (Bidelman et al., 2014; Bidelman & Lee, 2015). Future work should aim to address the influence of stimulus familiarity via direct (short-term) perceptual training (Pavlik et al., 2013). Additionally, it would be of interest to examine possible interactions between attentional load and/or stimulus familiarity on the neural representations of CP with long-term music (e.g., Bidelman et al., 2014) or second language experience (e.g., Bidelman & Lee, 2015; Zhang et al., 2010).

Conclusions

In sum, we infer that the neural correlates of CP and robust behavioral identification require active engagement with stronger categorical neural coding being expressed during states of attention, providing support to prior research (Bidelman et al., 2013). Further, CP is contingent upon stimulus familiarity such that listeners who lack the necessary stored auditory information cannot make a comparative judgment regarding two or more contrasting stimuli. Behavioral responses are significantly predicted by changes in neural correlates of CP during active engagement only. Our results ultimately support the claims that the expression and saliency of CP is contingent upon attention and stored auditory information.

Acknowledgments

This work was supported by a grant from the GRAMMY® Foundation awarded to G.M.B.

References

- Alain, C., Roye, A., & Salloum, C. (2014). Effects of age-related hearing loss and background noise on neuromagnetic activity from auditory cortex. *Frontiers in Systems Neuroscience*, 8, 8.
- Alain, C., Snyder, J.S., He, Y., & Reinke, K.S. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cerebral Cortex*, 17, 1074–1084.
- Babyak, M., A. (2004) What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, 66,411-421.
- Bidelman, G. M. (2015). Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. *Journal of Neuroscience Methods*, 241, 94–100.
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, 23(2), 425-434.
- Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, 1355, 112–125. doi:10.1016/j.brainres.2010.07.100
- Bidelman, G. M., & Lee, C.-C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *NeuroImage*. doi:10.1016/j.neuroimage.2015.06.087
- Bidelman, G. M., Moreno, S., & Alain, C. (2013). 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. (2014). Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiology of Aging*, 35(11), 2526-2540.
- Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, 40, 2662–73. doi:10.1111/ejn.12627
- Bomba, M. D., Choly, D., & Pang, E. W. (2011). Phoneme discrimination and mismatch negativity in English and Japanese speakers. *Neuroreport*, 22, 479-83.
- Brattico, E., Pallesen, K. J., Varyagina, O., Bailey, C., Anourova, I., Jarvenpaa, M., Eerola, T., & Tervaniemi, M. (2008). Neural discrimination of nonprototypical chords in music experts and laymen: An MEG study. *Journal of Cognitive Neuroscience*, 21(11), 2230-2244.

- Burns, E. M., & Ward, W. D. (1978). Categorical perception – phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, *63*(2), 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. *Nature Neuroscience*, *13*, 1428–1432
- Chartrand, J.-P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, *405*, 164–167. doi:10.1016/j.neulet.2006.06.053
- Cooper, A., & Wang, Y. (2012) The influence of linguistic and musical experience on Cantonese word learning. *Journal of the Acoustic Society of America*, *131*, 4756–4769.
- Cutting, J. E., & Rosner, B. S. (1974). Categories and boundaries in speech and music. *Perception & Psychophysics*, *16*, 564-571.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science*, *171*(3968), 303–306.
- Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults, *NeuroReport*, *8*, 919–924
- Garinis, A. C., & Cone-Wesson, B. K. (2007). Effects of stimulus level on cortical auditory event-related potentials evoked by speech. *Journal of the American Academy of Audiology*. doi:10.3766/jaaa.18.2.3
- Habib, M., & Besson, M. (2009). What do Music Training and Musical Experience Teach Us about Brain Plasticity? *Music Perception*, *26*(3), 279–286.
- Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual, *Proceedings of the National Academy of Sciences*, *95*, 781–787.
- 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
- Kappenman, E. S., & Luck, S. J. (2011). *The Oxford Handbook of Event-Related Potential Components ERP Components and Selective Attention*. doi:10.1093/oxfordhb/9780195374148.013.0144
- Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, 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. *International Journal of Psychophysiology*, *67*, 131-143.

- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606-608.
- Luck, S.J. (2005). *An Introduction to the Event-related Potential Technique*. MIT press, Cambridge, MA.
- Luck, S. J., Heinze, H. J., Mangun, G. R., & Hillyard, S. A. (1990). Visual event- related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. *Electroencephalography and Clinical Neurophysiology*, 75, 528–542.
- 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. *Cognitive, Affective, and Behavioral Neuroscience*, 2, 89-108.
- McClelland, J. L., Fiez, J.A., & McCandliss, B. D. (2002). Teaching the /r/-/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiology and Behavior*, 77, 657-662.
- McMullen, E., & Saffran, J. R. (2014). Music and Language: A Developmental Comparison. *Music Perception: An Interdisciplinary Journal*, 21(3), 289–311.
- Naatanen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, 13, 210-288.
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1-21.
- Näätänen, R., Gaillard, A.W.K., & Mäntysalo, S. (1978). Early selective attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42, 313–329.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour-Luhtanen, M., Huotilainen, M., Iivonen, A., Vainio, ... Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432–434.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*. doi:10.1016/j.clinph.2007.04.026
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24, 375–425.

- Nash, A., Sharma A, Martin, K., & Biever A. (2008). *Clinical applications of the P1 cortical auditory evoked potential (CAEP) biomarker*. A sound foundation through early amplification: proceedings of a Fourth International Conference. Stafa, Switzerland: Phonak, AG.
- 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. *Clinical Neurophysiology*, 112, 713-719.
- Pantev, C., & Herholz, S. C. (2011). Plasticity of the human auditory cortex related to musical training. *Neuroscience and Biobehavioral Reviews*, 35(10), 2140–54.
doi:10.1016/j.neubiorev.2011.06.010
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L.E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, 392, 811–814.
- Pavlik, P., Hua, H., Williams, J., & Bidelman, G. M. (2013). Modeling the effect of spacing on musical interval training. Paper presented at the Proceedings of 6th International Conference on Educational Data Mining, Memphis, TN
- Picton, T.W., Alain, C., Otten, L., Ritter, W., & Achim, A., (2000). Mismatch negativity: different water in the same river. *Audiology and Neurotology*. 5, 111–139.
- Picton, T. W., Woods, D. L., Baribaeu-Braun, J., & Healy, T. M. G. (1977). Evoked potential audiometry. *Journal of Otolaryngology*, 6(2), 90-119.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253–260.
- Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, 25(1-2), 21–52.
- Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics*, 15(2), 285-290.
- Seppanen, M., Hamalainen, J., Pesonen, A.K., & Tervaniemi, M. (2012). Music training enhances rapid neural plasticity of n1 and p2 source activation for unattended sounds. *Frontiers in Human Neuroscience*, 6, 43.
- Shahin, A., Bosnyak, D., Trainor, L., & Roberts, L. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*, 23, 5545–5552.
- Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America*, 106, 1078–83.

- Sharma, A., & Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America*, *107*, 2697–2703.
- Sussman, E., Winkler, I., & Wang, W. (2003). MMN and attention: Competition for deviance detection. *Psychophysiology*. doi:10.1111/1469-8986.00045
- Sussman, E. S. (2007). A New View on the MMN and Attention Debate The Role of Context in Processing Auditory Events. *MMN and Attention Journal of Psychophysiology*, *21*(3). doi:10.1027/0269-8803.21.3.xxx
- Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear and Hearing*, *22*, 79–90.
- Virtala, P., Putkinen, V., Huotilainen, M., Makkonen, T., & Tervaniemi, M. (2012). Musical training facilitates the neural discrimination of major vs. minor chords in 13-year-old children. *Psychophysiology*, *49*, 1125-1132. doi: 10.1111/j.1469-8986.2012.01386.x.
- Winkler, I. (2007). Interpreting the Mismatch Negativity. *Interpreting MMN Journal of Psychophysiology*, *21*. Doi:10.1027/0269-8803.21.3.147
- Winkler, I., Kujala, T., Titinen, H., Sivonen, P., Alku, P., Lehtokoski, A., ... Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*, 638-642.
- Wood, C.C., Goff, W.R., & Day, R.S. (1971). Auditory evoked potentials during speech perception. *Science*, *173*, 1248–1251.
- Ylinen S, Shestakova A, Huotilainen M, Alku P, & Naatanen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research*, *1072*(1): 175-185.
- Zendel, B. R., & Alain, C. (2013). The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *Journal of Cognitive Neuroscience*, *25*(4), 503-516.
- Zhang, L., Xi, J., Wu, H., Shu, H., & Li, P., (2012). Electrophysiological evidence of categorical perception of Chinese lexical tones in attentive condition. *Neuroreport* *23*, 35–39.
- 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.

Appendix A

ANOVA results for latency measures for the individual waves (P1, N1, P2) of the ERP. No main (condition, stimulus category) or interaction (condition x category) effects were observed for any of the latency measures in either the speech or music domain.

P1 latency music

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	1	9	0.04	0.8447
cat	1	9	1.08	0.3249
condition cat	1	9	0.43	0.5279

P1 latency speech

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	1	9	1.26	0.2904
cat	1	9	0.53	0.4866
condition cat	1	9	4.37	0.0662

N1 latency music

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	1	9	0.32	0.5826
cat	1	9	0.14	0.7138
condition cat	1	9	0.04	0.8413

N1 latency speech

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	1	9	0.01	0.9394
cat	1	9	0.22	0.6538
condition*cat	1	9	0.01	0.9144

P2 latency music

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	1	9	0.87	0.3754
cat	1	9	1.50	0.2516
condition*cat	1	9	0.32	0.5843

P2 latency speech

Type III Tests of Fixed Effects				
Effect	um DF	Den DF	F Value	Pr > F
condition	1	9	0.00	0.9685
cat	1	9	0.47	0.5093
condition*cat	1	9	1.23	0.2956

Appendix B

Below is a copy of the questionnaire administered to participants prior to the experiment

GENERAL INFORMATION:

1. Participant #: _____
2. Gender: Male Female
3. Date of birth: ____ (month)/ ____ (day)/ ____ (year)
4. Current Age: _____
5. Birth Place: _____
6. How long have you been in USA: ____ years ____ months
7. As far as you know, do you have normal hearing? Yes No
8. Are you left or right handed? Left Right
9. Have you ever had or been hospitalized for any neurological or psychiatric problems or a brain injury (if yes, please describe)?
Yes No
10. Do you drink? (If yes, please indicate how many times per week/month) Yes No
11. Do you use recreational drugs? (If yes, please indicate how many times per week/month) Yes No
12. Do you smoke? Yes No
13. Do you wear corrective lenses? Yes No
14. What is your general health? Excellent Good Fair Poor
15. Have you ever failed or skipped a grade? Yes No
16. Are you currently taking any medications? (If yes, please list) Yes No
17. Indicate your parent(s) highest level of education according to the 6 point scale: 1) some high school; (2) high school diploma or GED; (3) some college, vocational degree, or associate's degree; (4) 4-year college degree (e.g., BA, BS); (5) master's degree (e.g., MA, MS, MBA); (6) doctoral degree (e.g., PhD, MD, JD, EdD, ThD).

EDUCATIONAL BACKGROUND:

1. Degrees held (circle all that apply): none BA BS MA MS PhD MD other _____

(name of institution) (country) (year earned)

(major) (languages of instruction)

2. Total number of years of formal education (elementary school to present) = _____

LANGUAGE BACKGROUND:

1. What is your native language? (If you grew up with more than one language, please specify):

2. Other than English, what languages can you speak fluently?: _____

3. List all foreign languages you know in order of most proficient to least proficient. Rate your ability on the following aspects in each language. Please rate according to the following scale (write down the number in the table):

very poor poor fair functional good very good native-like
 1 _____ 2 _____ 3 _____ 4 _____ 5 _____ 6 _____ 7 _____

Language	Reading proficiency	Writing proficiency	Speaking fluency	Listening ability

4. Estimate, in terms of percentages, how often you use your native language and other languages per day (in all daily activities combined):

Native language _____%

Second language _____%

Other languages _____% (specify: _____)

(Total should equal 100%)

5. What is the age at which you started to learn your second language in the following situations (write age next to any situation that applies)?

At home _____

In school _____

After arriving in the second language speaking country _____

6. How did you learn your second language? (circle all that apply)

(Mainly Mostly Occasionally) through formal classroom instruction.

(Mainly Mostly Occasionally) through interacting with people.

A mixture of both, but (More classroom More interaction Equally both).

Other (**specify:** _____).

7. At present, which language do you prefer to speak in? (Circle one):

a. English b. Cantonese c. Regional dialects (s) d. Other _____

8. How would you rate your ability to communicate orally in English?:

Excellent Good Average Fair Poor

9. What language do you usually speak to your mother/father at home? (If not applicable for any reason, write N/A)

10. What language(s) do your parents usually speak to each other at home? (If not applicable for any reason, write N/A)

11. Write down the name of the language in which you received instruction in school, for each schooling level:

Primary/Elementary School _____

Secondary/Middle School _____

High School _____

College/University _____

12. In which languages do you usually:

Add, multiply, and do simple arithmetic? _____

Dream? _____

Express anger or affection? _____

13. In which language (among your best two languages) do you feel you usually do better? Write the name of the language under each condition.

	At home	At work
Reading	_____	_____
Writing	_____	_____
Speaking	_____	_____
Understanding	_____	_____

14. If you have taken a standardized test of proficiency for languages other than your native language (e.g., TOEFL or Test of English as a Foreign Language), please indicate the scores you received for each.

Language	Scores	Name of the Test
_____	_____	_____
_____	_____	_____
_____	_____	_____

MUSICAL BACKGROUND:

1. Have you ever had private or group (ensemble) music lessons (e.g., private or group music lessons)?
 Yes No

2. If you answered “yes” to question 1, please state the type of lessons you had, the instrument studied, and the year(s) you were enrolled in the lessons and/or class, the age you started playing, your proficiency (1=not proficient. 10=professional), and the style/genre (e.g., jazz piano; classical voice) of the music studied. Please be as specific as possible. Please list your primary instrument first followed by any secondary instruments:

Type (group/private)	Instrument (e.g., violin)	Dates (e.g., 1991-1992)	Age Started	Proficiency	Style
_____	_____	_____	_____	_____	_____
_____	_____	_____	_____	_____	_____

3. Please list the number of hours per week (on average) that you practice or play your instrument (within the past 24 months). For previous instruments list the hours per week you used to practice.

Instrument	Hours of practice and/or lessons per Week
-------------------	--

_____	_____
_____	_____
_____	_____

4. Please rate your musical ability on the scale below (0 = no musical ability at all, 3.5 = some musical ability, 7 = professional musician)

0 1 2 3 4 5 6 7

5. Can you read music (0 = not at all, 3.5 = somewhat, 7 =fluently)?

0 1 2 3 4 5 6 7

6. If you have not had lessons (group/private) within the past 5 years have you continued to practice your instrument independently since? Yes No

7. Given a starting pitch, rate your ability to transcribe a simple melody by ear (i.e. notate what is heard).

(0=complete guess, 3.5 = mostly correct with a few mistakes in rhythm or pitch, 7=perfect transcription)

0 1 2 3 4 5 6 7

8. Do you have an experience with music theory or ear training (if yes, please describe):

9. Do you or does anyone in your family have perfect pitch (i.e., name notes by ear)?

I do Family member _____ No one that I am aware of (including myself)

10. Do you have/are working on a degree in music (if yes, please provide details)?

Yes _____ No

11. Have you ever taught music lessons? (if yes, please describe; dates, etc.)?

Yes _____ No

12. Please state any additional information about your music background in the space below.

Edinburgh Handedness Inventory

Please indicate your preferences in the use of hands in the following activities by putting a “+” in the appropriate column. Where the preference is so strong that you would never try to use the other hand unless absolutely forced to, put “++”. If, in any case, you are really indifferent put a “+” in both columns. Some of the activities require both hands. In these cases the part of the task, or object, for which hand-preference is wanted is indicated in brackets. Please try to answer all the questions, and only leave a blank if you have no experience at all or the object or task.

		L	R
1	Writing		
2	Drawing		
3	Throwing		
4	Scissors		
5	Comb		
6	Toothbrush		
7	Knife (without fork)		
8	Spoon		
9	Hammer		
10	Screwdriver		
11	Tennis or badminton racket		
12	Chopsticks		
13	Baseball bat (upper hand)		
14	Golf Club (lower hand)		
15	Broom (upper hand)		
16	Shooting basketball		
17	Striking match (match)		
18	Opening jar		
19	Dealing cards (card being dealt)		
20	Threading needle or thread according to which is moved		

Handed ness of Family Members (R=right handed; L=left-handed)

Father R L Mother R L

Sister₁ R L Sister₂ R L

Appendix C

Below is a copy of the approved Institutional Review Board Letter #2370

THE UNIVERSITY OF MEMPHIS

Institutional Review Board

To: Gavin Bidelman
gmbdlman@memphis.edu

From: Chair, Institutional Review Board
For the Protection of Human Subjects
irb@memphis.edu

Subject: Neural correlates of complex auditory perception (#2370)

Approval Date: October 3, 2012

This is to notify you of the board approval of the above referenced protocol. This project was reviewed at the expedited level in accordance with all applicable statutes and regulations as well as ethical principles.

Approval of this project is given with the following obligations:

1. At the end of one year from the approval date, an approved renewal must be in effect to continue the project. If approval is not obtained, the human consent form is no longer valid and accrual of new subjects must stop.

2. When the project is finished or terminated, the attached form must be completed and sent to the board.
3. No change may be made in the approved protocol without board approval, except where necessary to eliminate apparent immediate hazards or threats to subjects. Such changes must be reported promptly to the board to obtain approval.
4. The stamped, approved human subjects consent form must be used unless your consent is electronic. Electronic consents may not be used after the approval expires. Photocopies of the form may be made.

This approval expires one year from the date above, and must be renewed prior to that date if the study is ongoing.

Chair, Institutional Review
Board The University of
Memphis