1-1-2018

DISSOCIABLE MECHANISMS OF CONCURRENT SPEECH IDENTIFICATION IN NOISE AT CORTICAL AND SUBCORTICAL LEVELS.

Anusha Yellamsetty

DISSOCIABLE MECHANISMS OF CONCURRENT SPEECH IDENTIFICATION IN NOISE AT CORTICAL AND SUBCORTICAL LEVELS

by

Anusha Yellamsetty

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Communication Sciences and Disorders

The University of Memphis

August 2018

Dedication

I would like to dedicate my thesis to my beloved parents, grandparents and sisters.

Acknowledgements

Over the past four years I have received support and encouragement from a great number of individuals. I would like to express my sincere gratitude to my advisor, Dr. Gavin M Bidelman, for the continuous support of my Ph.D dissertation, for his patience, motivation and immense knowledge. His guidance helped me in all the time of research and writing of this dissertation. I could not have imagined having a better advisor and mentor for my Ph.D study. My sincere thanks for tolerating my endless questions.

Dr. Shaum Bhagat, I would like to express my deepest appreciation for seeing potential in me that I didn't know existed. Thank you for your kindness and encouragement.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Eugene Buder, Prof. George Relyea and Dr. Saradha Ananthakrishnan, for their encouragement, insightful comments and questions which gave me incentive to widen my research from various perspectives. My sincere thanks also goes to Dr. Lisa Lucks Mendel, without her precious support it would not be possible to complete my doctoral program. In, particular, I am grateful to Dr. Madhuri Gore for enlightening me the first glance of research.

I thank my fellow labmates in for the stimulating discussions, and Ph.D colleagues/friends for all the fun we have had in the last four years. I thank my friends: Guru, SivaRam, Chandan, Lipika, JISNAR for listening, supporting and cheering me up.

I want to take this opportunity to thank all the professors and staff in CSD for their support throughout my doctoral program.

Last, but certainly not least, I would like to thank my family: my parents, grandparents, sisters, brother-in-law and my lovely niece. They have assisted me in innumerable ways. Amma

and sissies, thank you for tolerating my episodes of lunacy, love you. None of this would be possible without my people.

This research was supported by the Institute for Intelligent Systems Student Dissertation Grant Program at the University of Memphis.

Preface

Chapter 2 was published as a manuscript in *Hearing Research.* Yellamsetty, A., & Bidelman, G. M. (2018). Low-and high-frequency cortical brain oscillations reflect dissociable mechanisms of concurrent speech segregation in noise. *Hearing research*, *361*, 92-102.

Chapter 3 is in preparation for submission as a manuscript to *"Neuroscience"* Subcortical processing of concurrent speech mixtures as revealed by frequency-following responses (FFRs).

Abstract

Yellamsetty, Anusha. PhD. The University of Memphis. August 2018. Dissociable mechanisms of concurrent speech segregation in noise at cortical and subcortical levels. Primary Advisor: Gavin M Bidelman, PhD.

When two vowels with different fundamental frequencies (F0s) are presented concurrently, listeners often hear two voices producing different vowels on different pitches. Parsing of this simultaneous speech can also be affected by the signal-to-noise ratio (SNR) in the auditory scene. The extraction and interaction of F0 and SNR cues may occur at multiple levels of the auditory system. The major aims of this dissertation are to elucidate the neural mechanisms and time course of concurrent speech perception in clean and in degraded listening conditions and its behavioral correlates. In two complementary experiments, electrical brain activity (EEG) was recorded at *cortical* (EEG Study #1) and *subcortical* (FFR Study #2) levels while participants heard double-vowel stimuli whose fundamental frequencies (F0s) differed by zero and four semitones (STs) presented in either clean or noise degraded (+5 dB SNR) conditions. Behaviorally, listeners were more accurate in identifying both vowels for larger F0 separations (i.e., 4ST; with pitch cues), and this F0-benefit was more pronounced at more favorable SNRs. Time-frequency analysis of *cortical* EEG oscillations (i.e., "brain rhythms") revealed a dynamic time course for concurrent speech processing that depended on both extrinsic (SNR) and intrinsic (pitch) acoustic factors. Early high frequency activity reflected pre-perceptual encoding of acoustic features (~200 ms) and the quality (i.e., SNR) of the speech signal (~250-350ms), whereas later-evolving low-frequency rhythms (~400-500ms) reflected post-perceptual, cognitive operations that covaried with listening effort and task demands. Analysis of *subcortical* responses indicated that while FFRs provided a high-fidelity representation of double vowel stimuli and the spectro-temporal nonlinear properties of the

peripheral auditory system. FFR activity largely reflected the neural encoding of stimulus features (exogenous coding) rather than perceptual outcomes, but timbre (F1) could predict the speed in noise conditions. Taken together, results of this dissertation suggest that subcortical auditory processing reflects mostly exogenous (acoustic) feature encoding in stark contrast to cortical activity, which reflects perceptual and cognitive aspects of concurrent speech perception. By studying multiple brain indices underlying an identical task, these studies provide a more comprehensive window into the hierarchy of brain mechanisms and time-course of concurrent speech processing.

Table of Contents

Chapter

1 General Introduction 1

2 Low- and high-frequency cortical brain oscillations reflect dissociable mechanisms of concurrent speech segregation in noise 6

Abstract 6

Introduction 7

Method 11

Subjects 11

General Speech-in-noise recognition task 12

Electrophysiological procedures 12

Double vowel Stimuli 12

EEG data recording and preprocessing 14

EEG time-frequency analysis 15

Behavioral data analysis 17

Identification accuracy and the “F0 benefit” 17

Reaction time (RTs) 18

Statistical analysis 18

Results 19

Behavioral data 19

Neural oscillatory responses during double-vowel coding 21

Brain-behavior relationships 24

Discussion 26

Effects of SNR and F0 cues on behavioral concurrent vowel 27
Cortical oscillations reveal mechanistic differences in concurrent speech segregation divisible by frequency band 28
On the additivity vs. interactions of cues for concurrent sound segregation 31
Directions for future work 33
Conclusions 33
3 Subcortical processing of concurrent speech mixtures as revealed by frequency-following responses (FFRs) 35
Abstract 35
Introduction 36
Method 40
Participants 40
Stimulus and Behavioral task 41
Double vowel stimuli 41
Behavioral double-vowel identification task. 42
FFR data recording and preprocessing 42
FFR analysis 43
Behavioral data analysis 44
Identification accuracy and the “F0 benefit” 44
Reaction time (RTs) 44
Statistical analysis 44
Results 45

Behavioral data 45

FFR responses to single and double vowels 46

Brain-behavior relationships 50

Discussion 52

Effects of SNR and F0 cues on behavioral concurrent vowel identification 52

Subcortical encoding of single vs. double vowels 53

Subcortical correlates of double vowel perception 54

4. General Discussion 59

Future directions 64

References 65

Appendix 76

List of Figures

Figure		Page
1	Behavioral responses for segregating double-vowel stimuli	20
2	Neural oscillatory responses to concurrent speech sounds are modulated by SNR and the presence/absence of pitch cues	22
3	Band-specific time-course during double-vowel segregation	23
4	Band-specific mean spectral peak amplitude across conditions	24
5	Brain-behavior regression underlying double-vowel segregation	26
6	Behavioral responses for segregating double-vowel stimuli	46
7	Brainstem FFR to double vowel mixture	48
8	Additive noise vs. speech-on-speech masking effects at 0ST (A) and 4ST (B).	49
9	Brain-behavior correlations underlying double-vowel perception	50
10	FFRs are modulated by stimulus salience rather than perceptual dominancy	51
11	Distribution of the Brain-behavior correlates at cortical and subcortical levels for double-vowel perception	61
12	The hierarchical neural to behavioral correlates that are driving the identification of the concurrent double vowel stimulus.	63

Chapter 1

GENERAL INTRODUCTION

For speech comprehension in noisy environments (e.g., cocktail parties), listeners must parse an acoustic mixture into groups sound elements coming from one source (i.e., one talker) and segregate these from other sources (i.e., other talker). The process of auditory streaming is thought to rely on several acoustic principles including the degree of (in)harmonicity (Alain, Arnott, & Picton, 2001; Bidelman & Alain, 2015a), temporal coherence/(a)synchrony (Van Noorden, 1975), spectral content, and spatial configurations between multiple auditory objects (for reviews, see Bidet-Caulet & Bertrand, 2009; Bregman, 1990a; Oxenham, 2008; Shamma, Elhilali, & Micheyl, 2011). In particular, differences in the fundamental frequency (F0) between two or more sounds (i.e., pitch cues) represents one of the most robust acoustic factors for perceptual segregation. Auditory stimuli containing the same F0 are perceived as a single perceptual object whereas multiple F0s tend to promote hearing multiple sources. For instance, using synthetic double-vowel stimuli in a concurrent speech identification task, studies have shown that accuracy of identifying both vowels improves by 18% with increasing pitch differences between the vowels for F0 separations from 0 to about 4 semitones (STs) (Assmann & Summerfield, 1989a, 1990a, 1994a; de Cheveigné, Kawahara, Tsuzaki, & Aikawa, 1997). Research so far suggests that along with the F0 cues (Arehart, King, & McLean-Mudgett, 1997; Chintanpalli, Ahlstrom, & Dubno, 2016; Chintanpalli & Heinz, 2013a) listeners use additional acoustic cues to segregate speech such as spectral differences associated with formants (Ananthakrishna Chintanpalli & Heinz, 2013b), temporal envelope cues like harmonic interactions (Culling & Darwin, 1993), and spectral edges.

The segregation of speech and non-speech signals is thought to involve a multistage hierarchy of processing, whereby initial pre-attentive processes partition the sound waveform into distinct acoustic features (e.g. pitch, harmonicity) which is then acted upon by later, post-perceptual Gestalt principles (Koffka, 1935) [e.g., grouping by physical similarity, temporal proximity, good continuity (Bregman, 1990b)] and phonetic template matching (Alain, Reinke, He, Wang, & Lobaugh, 2005; Meddis & Hewitt, 1992a). Thus, the distributed neural network involves both subcortical and cortical brain regions (Alain, Reinke, McDonald, et al., 2005; Bidelman & Alain, 2015b; Dyson & Alain, 2004; Sinex, Sabes, & Li, 2002a).

Studies that directly examined the neural underpinnings of segregation of double vowel speech stimuli showed that neural encoding is not the same at different levels of the auditory pathway. The temporal discharge patterns and the spatial distribution auditory nerve (AN) fibers and cochlear CN contained sufficient information to identify both F0s (Keilson, Richards, Wyman, & Young, 1997; Alan R Palmer & Winter, 1992; AR Palmer, 1990c). Whereas at inferior colliculus (IC), neurons are tuned to spectral peaks (formants)(Carney, Li, & McDonough, 2015) and poorly represented F0 (Sinex, 2008; Sinex, Henderson, Li, & Chen, 2002; Sinex, Li, & Velenovsky, 2005; Sinex, Sabes, & Li, 2002b). Cortically, event-related brain potentials (ERPs) have mapped the time course of concurrent speech processing modulations in neural activity have been observed as early as ~150-200 ms, indicative of pre-attentive signal detection, with conscious identification of simultaneous speech occurring slightly later, ~350-400 ms post-stimulus onset (Alain, Arsenault, Garami, Bidelman, & Snyder, 2017; Alain, Reinke, He, et al., 2005; C. Alain, Snyder, He, & Reinke, 2007; Bidelman & Yellamsetty, 2017a; Reinke, He, Wang, & Alain, 2003).

One of the main factors affecting the parsing of simultaneous speech in the real world is signal-to-noise ratio (SNR). Additive noise tends to obscure less intense portions of the speech signal, preventing audible access to the salient speech cues normally exploited for comprehension (e.g., temporal envelope; Bidelman, 2016; Shannon et al., 1995; Swaminathan and Heinz, 2012). Successful perception of concurrently presented speech in noise is dependent on cognitive factors as well as sound processing at peripheral, subcortical and cortical levels, making it the most complex aspects of human communication (Kujala & Brattico, 2009; Shinn-Cunningham & Best, 2008). Studies so far have shed light on evoked cortical activity underlying the neural encoding of concurrent speech and have focused on how listeners track dynamic F0 information, and how the pitch cues aid the monitoring of auditory sources (Assmann, 1996) and improve speech perception in noise (Bidelman and Krishnan, 2010; Macdonald et al., 2010; Nabelek et al., 1989). ERPs cannot speak to the potential connection between cognitively driven non-phase locked *induced* brain rhythms and concurrent speech. These intrinsic brain rhythms are temporally jittered and are washed away by traditional time-locked ERP averaging. Studying brain rhythms would give us a platform to understand the mechanisms of perception and cognitive processing involved in concurrent speech identification task.

On the other hand, in studies at subcortical level, investigators typically manipulate the amount of acoustic information in the stimulus (e.g., SNR) and observe parallel changes in neural responses for isolated speech sounds (e.g., vowel, stop consonants). In such experimental designs, modulations in the evoked responses and human behavior both covary with the acoustic properties of the signal. This confounding of variables is further obscured if changes in the subcortical pre-attentive neural activity reflect a true correlate of the auditory percept or merely reflect properties of the stimulus itself. This distinction is important as recent Frequency

Following Responses (FFR) studies have shown the dissociation of acoustics from the actual percept suggesting that FFRs may not reflect a true neural correlate of the auditory percept but rather reflects more exogenous stimulus properties (Bidelman, 2017b; Bidelman, Moreno, & Alain, 2013b; Gockel, Carlyon, Mehta, & Plack, 2011). We are not aware of any studies examining the dissociation of acoustics from the actual percept, and how composite noise and pitch information affect the parsing of simultaneous speech across the auditory pathway.

To this end, we recorded EEG at subcortical (pre-attentive) and cortical (post-attentive) levels for a concurrent speech stimulus with two major aims (1) To illustrate the hierarchy of the connectivity between neuro-electric brain streams elicited by concurrent speech identification; and (2) To elucidate the neural mechanisms and time course of concurrent speech identification in clean and degraded listening conditions. Subcortically, FFRs allow us to estimate how salient properties of speech spectra (e.g., F0s or formants of concurrent vowels) are transcribed by the *human* auditory nervous system at early, pre-attentive stages of the processing hierarchy. Cortically, new time-frequency analysis of the EEG provided novel insight into the correspondence between cortical brain rhythms and speech perception and how listeners exploit pitch and SNR cues for successful identification. We hypothesized, that the spectral components of FFRs reflect the encoding of non-linear interactions between the two concurrent vowels. Additionally, FFRs would show reduced amplitudes with noise and correlate with behavioral identification scores, offering an objective, subcortical correlates of concurrent speech perception. With cortical neural rhythms, we expected that early modulations in higher frequency bands of the EEG (i.e., γ-band) would be sensitive to the acoustic features of stimuli and the quality of speech representations. Alternatively, the lower frequency bands of oscillation (i.e., θ-

band) would reflect more domain general, or internal operations related to the perceptual segregation process and task demands (such as attention, listening effort, or memory demands).

Chapter 2

LOW- AND HIGH-FREQUENCY CORTICAL BRAIN OSCILLATIONS REFLECT DISSOCIABLE MECHANISMS OF CONCURRENT SPEECH SEGREGATION IN NOISE

Abstract

Parsing simultaneous speech requires listeners use pitch-guided segregation which can be affected by the signal-to-noise ratio (SNR) in the auditory scene. The interaction of these two cues may occur at multiple levels within the cortex. The aims of the current study were to assess the correspondence between oscillatory brain rhythms and determine how listeners exploit pitch and SNR cues to successfully segregate concurrent speech. We recorded electrical brain activity while participants heard double-vowel stimuli whose fundamental frequencies (F0s) differed by zero or four semitones (STs) presented in either clean or noise-degraded (+5dB SNR) conditions. We found that behavioral identification was more accurate for vowel mixtures with larger pitch separations but F0 benefit interacted with noise. Time-frequency analysis decomposed the EEG into different spectrotemporal frequency bands. Low-frequency (θ, β) responses were elevated when speech did not contain pitch cues (0ST>4ST) or was noisy, suggesting a correlate of increased listening effort and/or memory demands. Contrastively, γ power modulations were observed for changes in both pitch (0ST>4ST) and SNR (clean>noise), suggesting high-frequency bands carry information related to acoustic features and the quality of speech representations. Brain-behavior associations corroborated these effects; modulations in low-frequency rhythms predicted the speed of listeners' perceptual decisions with higher bands predicting identification accuracy. Results demonstrate that neural oscillations reflect both

automatic (pre-perceptual) and controlled (post-perceptual) mechanisms of speech processing that are largely divisible into high- and low-frequency bands of human brain rhythms.

Keywords: EEG; time-frequency analysis; double-vowel segregation; F0-benefit; speech-in-noise perception

INTRODUCTION

In normal auditory scenes (e.g., cocktail parties), listeners must parse acoustic mixtures to extract the intended message of a target, a process known as source segregation. Previous studies have suggested that fundamental frequency (F0) (i.e., pitch) differences provide a robust cue for identifying the constituents of concurrent speech. For instance, using synthetic double-vowel stimuli in a concurrent speech identification task, studies have shown that accuracy of identifying both vowels improves with increasing pitch differences between the vowels for F0 separations from 0 to about 4 semitones (STs) (Assmann & Summerfield, 1989b, 1990b, 1994b; de Cheveigné et al., 1997). This improvement has been referred to as the "F0 benefit" (Arehart et al., 1997; Chintanpalli et al., 2016; Chintanpalli & Heinz, 2013a). Thus, psychophysical research from the past several decades confirms that human listeners exploit F0 (pitch) differences to segregate concurrent speech.

Neural responses to concurrent speech and non-speech sounds have been measured at various levels of the auditory system including single-unit recordings in animals (AR Palmer, 1990c; Portfors & Sinex, 2005; Sinex, Guzik, Li, & Sabes, 2003; Snyder & Sinex, 2002) and in human, via evoked potentials (Alain, Reinke, He, et al., 2005; Bidelman, 2017a; Bidelman & Alain, 2015b; Dyson & Alain, 2004) and fMRI (Arnott, Grady, Hevenor, Graham, & Alain, 2005). The segregation of complex signals is thought to involve a multistage hierarchy of processing, whereby initial pre-attentive processes partition the sound waveform into distinct

acoustic features (e.g., pitch, harmonicity) which is then acted upon by later, post-perceptual Gestalt principles (Koffka, 1935) [e.g., grouping by physical similarity, temporal proximity, good continuity (Bregman, 1990b)] and phonetic template matching (Alain, Reinke, He, et al., 2005; R. Meddis & Hewitt, 1992b).

In humans, the neural correlates of concurrent speech segregation have been most readily studied using event-related brain potentials (ERPs). Modulations in ERP amplitude/latency provide an index of the timing and level of processing for emergent mechanisms of speech segregation. Mapping the time course of concurrent speech processing, modulations in neural activity have been observed as early as ~150-200 ms, indicative of pre-attentive signal detection, with conscious identification of simultaneous speech occurring slightly later, ~350-400 ms post-stimulus onset (Alain et al., 2017; Alain, Reinke, He, et al., 2005; C. Alain et al., 2007; Bidelman & Yellamsetty, 2017b; Du et al., 2010; Reinke et al., 2003). Further perceptual learning studies have shown enhancements in the ERPs with successful learning in double vowel tasks in the form of an earlier and larger N1-P2 complex (enhanced sensory coding < 200 ms) coupled with larger slow wave activity (~ 400 ms), indicative of more effective cognitive processing/memory template matching (C. Alain et al., 2007; Reinke et al., 2003). Using brain-imaging methods (PET, fMRI), the spatial patterns of neural activation associated with speech processing have also been visualized in various regions of the auditory cortex (Giraud et al., 2004; Pulvermüller, 1999). For example, fMRI implicates a left thalamocortical network including thalamus, bilateral superior temporal gyrus and left anterior temporal lobe in successful double-vowel segregation (Alain, Reinke, He, et al., 2005).

One of the main factors affecting the parsing of simultaneous speech is signal-to-noise ratio. In real-world listening environments, successful recognition of noise-degraded speech is

thought to reflect a frontotemporal speech network involving a close interplay between primary auditory sensory areas and inferior frontal brain regions (Bidelman & Alain, 2015b; Bidelman & Howell, 2016; Binder, Liebenthal, Possing, Medler, & Ward, 2004; Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010). Consequently, dynamic F0 cues and noise (SNR) are likely to interact during the extraction of multiple auditory streams and occur relatively early (within few hundred milliseconds) in the neural hierarchy (Bidelman, 2017a; Bidelman & Yellamsetty, 2017b).

While prior studies have shed light on cortical activity underlying the neural encoding of concurrent speech, they cannot speak to how different frequency bands of the EEG (i.e., neural oscillations) relate to concurrent speech segregation. These frequency-specific "brain rhythms" become apparent only after averaging single-trial epochs in the spectral domain. The resulting neural spectrogram can be decomposed into various frequency bands which are thought to reflect local (high-frequency) and long-range (low -frequency) communication between different neural populations. Studies also suggest that various frequency ranges of the EEG may reflect different mechanisms of processing, including attention (Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008), navigation (Buzsáki, 2005), memory (Palva, Monto, Kulashekhar, & Palva, 2010; Sauseng, Klimesch, Gruber, & Birbaumer, 2008), motor planning (Donoghue, Sanes, Hatsopoulos, & Gaál, 1998), and speech-language comprehension (Doelling, Arnal, Ghitza, & Poeppel, 2014; Ghitza, 2011, 2013; Ghitza, Giraud, & Poeppel, 2013; Haarmann, Cameron, & Ruchkin, 2002; Shahin, Picton, & Miller, 2009). Although still debated, the general consensus is that lower frequency oscillations are associated with the perception, cognition, and action, whereas high-frequency bands are associated with stimulus transduction, encoding, and feature selection (von Stein & Sarnthein, 2000).

With regard to speech listening, different oscillatory activity may contribute to the neural coding of acoustic features in the speech signal or different internal cognitive operations related to the perceptual segregation process. Speech can be decomposed into different bands of time-varying modulations (i.e., slow-varying envelope vs. fast-varying fine structure) which are captured in the neural phase-locked activity of the scalp EEG (Bidelman, 2016b). Theoretical accounts of brain organization suggest that different time-varying units of the speech signal (e.g., envelope vs. fine structure; phoneme vs. sentential segments) might be "tagged" by different frequency ranges of neural oscillations that coordinate brain activity at multiple spatial and temporal scales across distant cortical regions. Of relevance to speech coding, delta band (< 3Hz) oscillations have been shown to reflect processing related to sequencing syllables and words embedded within phrases (Ghitza, 2011, 2012). Theta (θ: 4-8 Hz) band has been linked with syllable coding at the word level (Bastiaansen, Van Der Linden, Ter Keurs, Dijkstra, & Hagoort, 2005; Giraud & Poeppel, 2012; Goswami, 2011) and attention/arousal (Aftanas, Varlamov, Pavlov, Makhnev, & Reva, 2001; Paus et al., 1997). In contrast, beta (β: 15-30 Hz) band has been associated with the extraction of global phonetic features (Bidelman, 2015a, 2017a; Fujioka, Trainor, Large, & Ross, 2012; Ghitza, 2011), template matching (Bidelman, 2015a), lexical semantic memory access (Shahin et al., 2009), and perceptual binding in brain networks (Aissani, Martinerie, Yahia-Cherif, Paradis, & Lorenceau, 2014; Brovelli et al., 2004; von Stein & Sarnthein, 2000). Lastly, gamma (γ: > 50 Hz) band has been associated with detailed phonetic features (Goswami, 2011), short duration cues (Giraud & Poeppel, 2012; Zhou, Melloni, Poeppel, & Ding, 2016), local network synchronization (Giraud & Poeppel, 2012; Haenschel, Baldeweg, Croft, Whittington, & Gruzelier, 2000), perceptual object construction (Tallon-Baudry & Bertrand, 1999a), and experience-dependent enhancements in speech processing

(Bidelman, 2017a). Yet, the role of rhythmic neural oscillations in concurrent speech perception and how various frequency bands of the EEG relate to successful auditory scene analysis remains unclear.

In the present study, we aimed to further elucidate the neural mechanisms of concurrent speech segregation from the perspective of *oscillatory* brain activity. To this end, we recorded neuroelectric responses as listeners performed a double-vowel identification task during stimulus manipulations designed to promote or deny successful segregation (i.e., changes in F0 separation of vowels; with/without noise masking). New time-frequency analysis of the EEG provided novel insight into the correspondence between brain rhythms and speech perception and how listeners exploit pitch and SNR cues for successful segregation. Based on previous investigations on evoked (ERP) correlates of concurrent speech segregation (C. Alain et al., 2007; Bidelman & Yellamsetty, 2017b; Reinke et al., 2003) we expected early modulations in higher frequency bands of the EEG (e.g., γ-band) would be sensitive to changes in F0-pitch and the SNR of speech. This would be consistent with the hypothesis that high frequency oscillations tag information related to the acoustic features of stimuli and the quality of speech representations. Additionally, we hypothesized that lower bands of oscillation (e.g., θ-band) would reflect more domain general, internal operations related to the perceptual segregation process and task demands (e.g., attention, listening effort, memory demands).

METHODS

Subjects

Thirteen young adults (mean ± SD age: 26.1 ± 3.8 years; 10 females, 3 males) participated in the experiment. All had obtained a similar level of formal education (19.6± 2.8 years), were right handed (>43.2 laterality) (Oldfield, 1971), had normal hearing thresholds (i.e., ≤ 25 dB HL) at octave frequencies between 250 and 8000 Hz, and reported no history of

neuropsychiatric disorders. Each gave written informed consent in compliance with a protocol approved by the University of Memphis Institutional Review Board.

General speech-in-noise recognition task

We measured listeners' speech-in-noise (SIN) recognition using the standardized QuickSIN test (Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004). We have previously shown a strong correspondence between QuickSIN scores and speech ERPs (Bidelman & Howell, 2016), justifying the inclusion of this instrument. Participants heard two lists embedded in four-talker babble noise, each containing six sentences with five key words. Sentences were presented at 70 dB SPL using pre-recorded signal-to-noise ratios (SNRs) which decreased in 5 dB steps from 25 dB (easy) to 0 dB (difficult). After each presentation, participants repeated the sentence and the number of correct key words were scored. "SNR loss" (computed in dB) was determined by subtracting the total number of correctly recalled words from 25.5. This metric represents the SNR required to correctly identify 50% of the key words across the sentences (Killion et al., 2004). SNR loss was measured for two lists separately for the left and right ear. The four responses were then averaged to obtain a stable SIN recognition score for each participant.

Electrophysiological procedures

Double vowel stimuli

Speech stimuli were modeled after previous studies on concurrent double-vowel segregation (C. Alain et al., 2007; Assmann & Summerfield, 1989b, 1990b; Bidelman & Yellamsetty, 2017b). Synthetic, steady-state vowel tokens (/a/, /i/, and /u/) were created using a Klatt synthesizer (Klatt, 1980) implemented in MATLAB® 2015b (The MathWorks, Inc.). Each

token was 200 ms in duration including 10-ms $\cos^2$ onset/offset ramping. Vowel F0 and formant frequencies were held constant over the duration. F0 was either 100 or 125 Hz. Double-vowel stimuli were then created by combining single-vowel pairs. Each vowel pair had either identical (0 ST) or different F0s (4ST). That is, one vowel's F0 was set at 100 Hz while the other vowel had an F0 of 100 or 125 Hz so as to produce double-vowels with an F0 separation of either 0 or 4 semitones (STs). Each vowel was paired with every other vowel (except itself), resulting in a total of 6 unique double-vowel pairings (3 pairs x 2 F0 combinations). Double-vowels were presented in a clean and noise condition (separate blocks), in which stimuli were delivered concurrently with a backdrop of multi-talker noise babble (+5 dB SNR) (Bidelman & Howell, 2016; Nilsson, Soli, & Sullivan, 1994). SNR was manipulated by changing the level of the masker rather than the signal to ensure that SNR was not positively correlated with overall sound level (Bidelman & Howell, 2016; Binder et al., 2004). Babble was presented continuously to avoid time-locking it with the stimulus presentation. We chose continuous babble over other forms of acoustic inference (e.g., white noise) because it more closely mimics real-world listening situations and tends to have a larger effect on the auditory ERPs (Kozou et al., 2005).

Stimulus presentation was controlled by MATLAB routed to a TDT RP2 interface (Tucker-Davis Technologies). Speech stimuli were delivered binaurally at an intensity of 81 dB SPL through ER-2 insert earphones (Etymotic Research). During EEG recording, listeners heard 50 exemplars of each double-vowel combination and were asked to identity *both* vowels as quickly and accurately as possible on the keyboard. Feedback was not provided. The inter-stimulus interval was jittered randomly between 800 and 1000 ms (20-ms steps, rectangular distribution) to avoid rhythmic entrainment of the EEG (Luck, 2005, p. 168) and listeners anticipating subsequent trials. The next trial commenced following the listener's behavioral

response. The order of vowel pairs was randomized within and across participants; clean and noise conditions were run in separate blocks alternatively. A total of six blocks (3 clean, 3 noise) were completed, yielding 150 trials for each of the individual double-vowel conditions. Listeners were given 2-3 min breaks after each block (10-15 min after 3 blocks) as needed to avoid fatigue.

Prior to the experiment proper, we required that participants be able to identify single vowels in a practice run with >90% accuracy (e.g., C. Alain et al., 2007). This ensured their task performance would be mediated by *concurrent* sound segregation skills rather than isolated identification, *per se.*

EEG data recording and preprocessing

EEG recording procedures followed well-established protocols in our laboratory (Bidelman, 2015b; Bidelman & Howell, 2016; Bidelman & Yellamsetty, 2017b). Neuroelectric activity was recorded from 64 sintered Ag/AgCl electrodes at standard 10-10 locations around the scalp (Oostenveld & Praamstra, 2001). Contact impedances were maintained <5 kΩ throughout the duration of the experiment. EEGs were digitized using a sampling rate of 500 Hz (SynAmps RT amplifiers; Compumedics Neuroscan). Electrodes placed on the outer canthi of the eyes and the superior and inferior orbit were used to monitor ocular activity. The data were pre-processed by thresholding EEG amplitudes at ±100 μV. Ocular artifacts (saccades and blink artifacts) were then corrected in the continuous EEG using a principal component analysis (PCA) (Wallstrom, Kass, Miller, Cohn, & Fox, 2004). Data were visually inspected for bad channels and paroxysmal electrodes were interpolated from the adjacent four nearest neighbor channels (distance weighted). These procedures helped remove myogenic and other artifacts prior to time-frequency analysis that can affect the interpretation of oscillatory responses (Pope, Fitzgibbon, Lewis, Whitham, & Willoughby, 2009). During online acquisition, all electrodes were

referenced to an additional sensor placed ~1 cm posterior to Cz. Data were re-referenced off-line to a common average reference. EEGs were then epoched (-200-1000 ms), baseline-corrected to the pre-stimulus interval, and digitally filtered (1-100 Hz, zero-phase) for response visualization and time-frequency analysis. To obtain an adequate number of trials for analysis, we pooled responses to collapse across different vowel pairs. This yielded 450 trials per listener for the four conditions of interest [i.e., 2 SNRs (clean, noise) x 2 F0s (0 ST, 4 ST)]. The entire experimental protocol including behavioral and electrophysiological testing took ~2 hrs. to complete.

EEG time-frequency analysis

Evoked potential (ERP) results related to this dataset are reported in our companion paper (Bidelman & Yellamsetty, 2017b). New time-frequency analyses (applied here) were used to evaluate the correspondence between *rhythmic* brain oscillations and speech perception and how listeners exploit pitch and SNR cues for successful segregation.

From epoched EEGs, we computed time-frequency decompositions of single-trial data to assess frequency-specific changes in oscillatory neural power (Bidelman, 2015a, 2017a). For each trial epoch, the time-frequency map (i.e., spectrogram) was extracted using Mortlet wavelets as implemented in the MATLAB package Brainstorm (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011b). This resulted in an estimate of the power for each time-frequency point over the bandwidth (1-100 Hz; 1 Hz steps) and time course (-200 – 1000 ms) of each epoch window. Using the Mortlet basis function, spectral resolution decreased linearly with increasing frequency; the full width half maximum (FWHM) was ~1 Hz near DC and approached ~20 Hz at 60 Hz. Temporal resolution improved exponentially with increasing frequency; FWHM was ~ 3 sec near DC and ~50 ms at 60 Hz. Single-trial spectrograms were then averaged across trials to obtain time-frequency maps for each subject and stimulus condition (see Fig. 2). When power is

expressed relative to the baseline pre-stimulus interval (-200 – 0 ms), these spectrographic maps are known as event-related spectral perturbations (ERSPs) (Delorme & Makeig, 2004). ERSPs represent the increase/decrease in EEG spectral power relative to the baseline pre-stimulus period (in dB). They contain neural activity that is both time- and phase-locked to the eliciting stimulus (i.e., evoked activity) as well as non-phase-locked responses (i.e., induced oscillatory activity) generated by the ongoing stimulus presentation (Bidelman, 2015a, 2017a; Shahin et al., 2009; Trainor, Shahin, & Roberts, 2009). To reduce the dimensionality of the data, we restricted our analysis to the Fz electrode. This channel is ideal for measuring auditory evoked responses (Picton et al., 1999a) and time-frequency oscillations (Bidelman, 2015a, 2017a) to speech which are both maximal over frontocentral scalp locations. Moreover, scalp topographies of our data (pooled across subjects and conditions) confirmed that most band responses were strongest near frontocentral regions of the scalp (see Fig. 3). While we restrict subsequent analyses to Fz, it should be noted that in pilot testing, we also analyzed responses at different electrode clusters. However, results were qualitatively similar to those reported herein (data not shown).

To quantify frequency-specific changes in oscillatory power to concurrent speech, we extracted time courses from ERSP maps in five different bands. Band-specific waveforms were extracted by taking "slices" of the ERSP maps averaged across different frequency ranges: 5-7 Hz (θ), 8-12Hz (α), 15-29 Hz (β), 30-59 Hz (γ_{low}), and 60-90 Hz (γ_{high}). This resulted in a running time waveform within each prominent frequency band of the EEG, similar to an ERP. We then contrasted band-specific waveforms (i.e., clean vs. noise; 0 ST vs. 4 ST) to compare the neural encoding of double-vowel stimuli across the main factors of interest (i.e., SNR and pitch). We used a running permutation test (EEGLAB's statcond function; Delorme & Makeig, 2004) to determine the time points over which band activity differed between stimulus conditions

(p<0.05, N=1000 resamples). We required that segments persisted contiguously for ≥25 ms to be considered reliable and help control false positives (Chung & Bidelman, 2016; Guthrie & Buchwald, 1991).

This initial analysis revealed time segments where band-specific oscillations were modulated by our stimulus manipulations (i.e., SNR and pitch). To better quantify stimulus-related changes, we extracted peak power from the mid-point of the time segments showing significant differences in band activity: θ: 450 ms; β: 350 ms; $\gamma_{low/high}$: average of peak power at 25 and 175 ms (see Fig. 3). Grand average ERSP scalp topographies (pooled across stimulus conditions) are shown for each band in Fig. 3. Scalp maps confirmed that synchronized responses to speech mixtures were maximal over the frontocentral plane (Claude Alain, Snyder, He, & Reinke, 2006; TW Picton et al., 1999b).

Behavioral data analysis

Identification accuracy and the "F0 benefit"

Behavioral identification was analyzed as the percent of trials where *both* vowel sounds were identified correctly. For statistical analyses, %-correct scores were arcsine transformed to improve homogeneity of variance assumptions necessary for parametric statistics (Studebaker, 1985). Increasing the F0 between two vowels provides a pitch cue which leads to an improvement in accuracy identifying concurrent vowels (Assmann & Summerfield, 1990b; R. Meddis & Hewitt, 1992b)—an effect referred to as the "F0-benefit" (Arehart et al., 1997; Bidelman & Yellamsetty, 2017b; Chintanpalli & Heinz, 2013a). To provide a singular measure of double-vowel identification we calculated the F0-benefit for each listener, computed as the difference in performance (%-correct) between the 4ST and 0ST conditions. F0-benefit was

computed separately for clean and noise stimuli allowing us to compare the magnitude of F0 benefit in concurrent speech segregation with and without noise interference.

Reaction time (RTs)

Behavioral speech labeling speeds [i.e., reaction times (RTs)], were computed separately for each participant as the median response latency across trials for a given double-vowel condition. RTs were taken as the time lapse between the onset of the stimulus presentation and listeners' identification of both vowel sounds. Following our previous studies on the neural correlates of speech perception (e.g., Bidelman et al., 2013b; Bidelman & Walker, 2017), RTs shorter than 250 ms or exceeding 6000 ms were discarded as implausibly fast responses and lapses of attention, respectively.

Statistical analysis

Unless otherwise noted, two-way, mixed-model ANOVAs were conducted on all dependent variables (GLIMMIX Procedure, SAS® 9.4, SAS Institute, Inc.). Stimulus SNR (2 levels; clean, +5 dB noise) and semitones (2 levels; 0ST, 4ST) functioned as fixed effects; subjects served as a random factor. Tukey-Kramer multiple comparisons-controlled Type I error inflation. An *a priori* significance level was set at α=0.05.

To examine the degree to which neural responses predicted behavioral speech segregation, we performed weighted least square regression between listeners' band-specific amplitudes and (i) their accuracy, and RTs in the double-vowel task and (ii) QuickSIN scores. Robust bisquare fitting was achieved using "fitlm" in MATLAB. To arrive at a comparable and single measure to describe how neurophysiological responses distinguished speech using pitch cues, we derived a "*neural* F0 benefit," computed as the difference between each listener's 4ST and 0ST responses. As in behavioral F0 benefit, this neural analogue was computed separately

for the clean and noise conditions. We then regressed behavioral and neural F0 benefits to assess brain-behavior correspondences. We reasoned that listeners who experience larger changes in their neural encoding of speech with added pitch cues (i.e., stronger neural F0 benefit) would have larger behavioral gains in the double-vowel segregation from 0 to 4 ST (i.e., experience bigger perceptual F0 benefit).

RESULTS

Behavioral data

Behavioral speech identification accuracy and RTs for double-vowel segregation are shown in Figure 1 A. Listeners obtained near-ceiling performance (96.9 ± 1.4%) when identifying single vowels. In contrast, double-vowel identification was considerably more challenging; listeners' accuracy ranged from ~30 – 70% depending on the presence of noise and pitch cues. An ANOVA conducted on behavioral accuracy confirmed a significant SNR x F0 interaction [$F_{1, 12} = 5.78$, $p = 0.0332$], indicating that successful double-vowel identification depended on both the noise level and presence of F0 pitch cues. Post hoc contrasts revealed listeners showed a similar level of performance with and without noise for 0 ST vowels, those which did not contain pitch cues. Performance increased ~30% across the board with greater F0 separations (i.e., 4ST > 0ST). F0-benefit was larger for clean relative to +5 dB SNR speech [$t_{12} = 2.15$, $p = 0.026$ (one-tailed)], suggesting listeners made stronger use of pitch cues when segregating clean compared to acoustically impoverished speech.

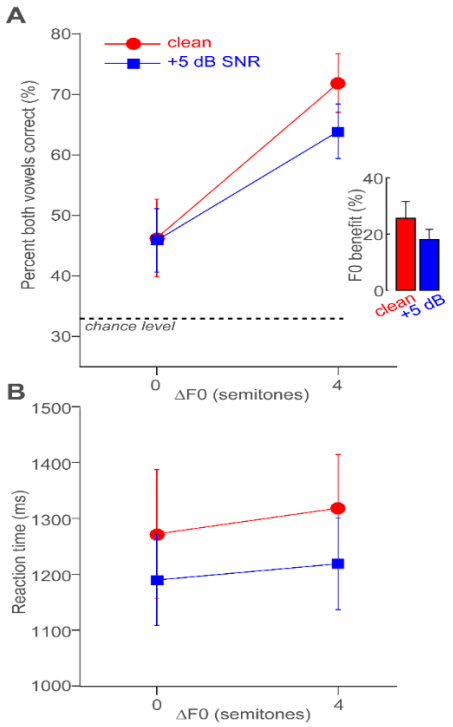

Figure 1. Behavioral responses for segregating double-vowel stimuli. (A) Accuracy for identifying both tokens of a two-vowel mixture. Performance is poorer when concurrent speech sounds contain the same F0 (0ST) and improve ~30% when vowels contain differing F0s (4ST). (*Insert*) Behavioral F0-benefit, defined as the improvement in %-accuracy from 0ST to 4ST, indexes the added benefit of pitch cues to speech segregation. F0-benefit is stronger for clean vs. noisy (+5 dB SNR) speech indicating that listeners are poorer at exploiting pitch cues when segregating acoustically-degraded signals. (B) Speed (i.e., RTs) for double-vowel segregation. Listeners are marginally faster at identifying speech in noise. However, faster RTs at the expense of poorer accuracy (panel A) suggests a time-accuracy tradeoff in double-vowel identification. Data reproduced from Bidelman and Yellamsetty (2017b). error bars = ±1 s.e.m.

Analysis of RTs revealed a marginal effect of SNR [$F_{1,\,12}$ = 4.11, p = 0.065]; listeners tended to be slower identifying clean compared to noisy speech (Fig. 1B). The slowing of RTs coupled with better %-identification for clean compared to noise-degraded speech is indicative of a time-accuracy tradeoff in concurrent sound segregation. Indeed, RTs and %-correct scores were highly correlated [r=0.46, p=0.006] such that more accurate identification corresponded with slower decisions.

Neural oscillatory responses during double-vowel coding

Grand average ERSP time-frequency maps are shown for each of the noise and ST conditions in Figure 2. Figure 3 shows time waveforms for the 5-7 Hz (θ), 8-12Hz (α), 15-29 Hz (β), 30-59 Hz (γ_{low}), and 60-90 Hz (γ_{high}) bands extracted from the spectrographic maps of Figure 2. Each reflects how different frequency oscillations in the EEG code double-vowel mixtures. Generally speaking, lower frequency bands including θ- and α-band showed sustained activity over the duration of the trial which appeared stronger for more difficult stimulus conditions (i.e., noisy speech and 0ST conditions). Compared to clean speech, β-band activity also appeared larger (more positive) ~400-500 ms after speech onset. Lastly, higher γ-band showed broadband transient activations that seem to tag the onset (see 25 ms) and offset (see 200 ms) of the evoking speech stimulus (cf. Ross, Schneider, Snyder, & Alain, 2010). These high γ-band events also appeared stronger for clean relative to noise-degraded speech and for 0ST vs. 4ST vowel mixtures. In terms of the overall time course of spectral responses, the strong modulations of high γ-band in clean and at 0ST are followed by negative modulation of β-band and sustained positive modulation of the θ- band. The directions of these band amplitude effects are reversed in the noise and 4 ST conditions.

Figure 3C shows the SNR x ST interaction waveforms. Interactions were confined to α- and β- bands, at early (~150-200 ms) time windows after stimulus onset. These early interactions replicate (are consistent with) the noise x pitch interactions observed in the N1-P2 time window of our previous ERP study on double-vowel coding (Bidelman & Yellamsetty, 2017b) and thus, were not explored further.

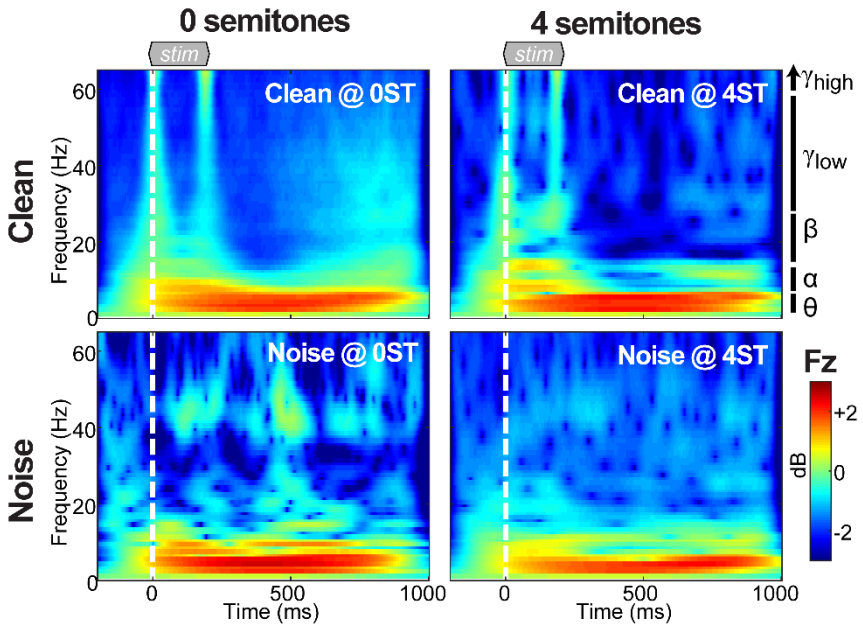

Figure 2. Neural oscillatory responses to concurrent speech sounds are modulated by SNR and the presence/absence of pitch cues. ERSP time-frequency maps (Fz channel) quantify both "evoked" and "rhythmic" changes in EEG power relative to the baseline period. Each panel represents the response to double-vowel stimuli with (4ST) or without (0ST) a difference in voice fundamental frequency for stimuli presented either in clean or +5 dB SNR of noise. Light gray regions above the spectrograms show the schematized stimulus. Dotted lines, stimulus onset (t=0).

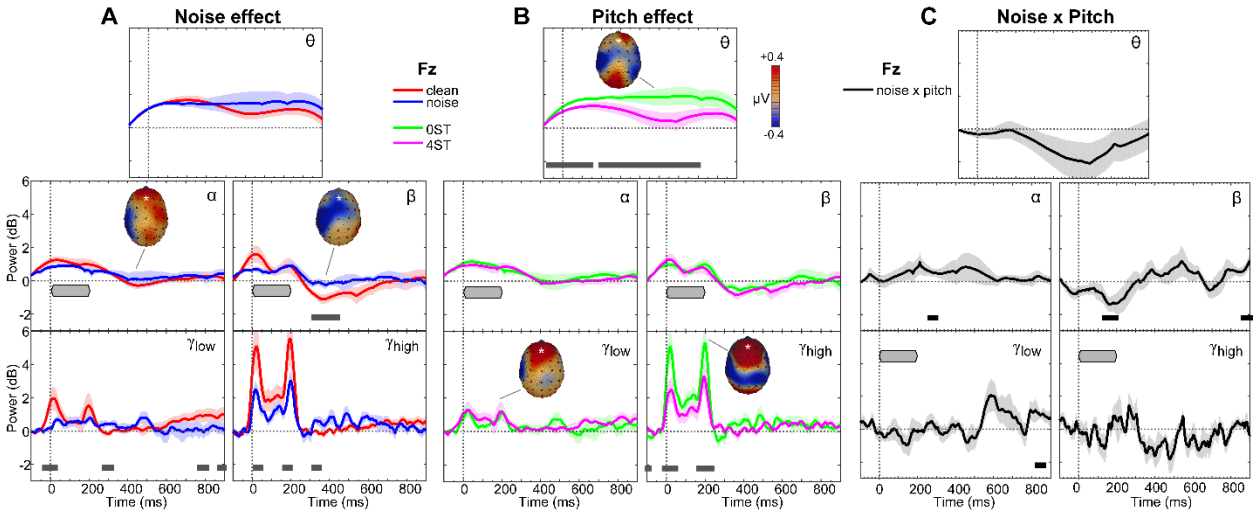

Figure 3. Band-specific time courses during double-vowel segregation. Shown here are response time courses for each frequency band of the EEG extracted from ERSP spectrograms and their interaction (see Fig. 2). Band waveforms contrast how noise SNR (**A**), F0 pitch (**B**), and their interaction (**C**; SNR x pitch) affect the neural encoding of double-vowel mixtures. A permutation test shows contiguous segments (≥25 ms duration) where spectral power differs between stimulus conditions (■ segments; $p < 0.05$, $N = 1000$ resamples). Modulations in β- and high γ-band distinguish clean from noise-degraded speech (β: clean < noise; γ_{high} = clean > noise). Contrastively, pitch cues are distinguished by modulations in the θ band (0ST > 4ST) and γ_{high} band (0ST > 4ST). Head maps (pooled across stimulus conditions and subjects) show the topographic distribution of each band across the scalp at time points where the band-specific effects are largest. * Fz electrode for subsequent analysis. Gray regions, schematized stimulus. Shading = ±1 s.e.m.

Next, we aimed to quantify changes in spectral band power due to each acoustic factor (SNR, STs). For each band time course for the two main effects (i.e., Fig. 3 and B), peak amplitudes were extracted from the temporal center of segments showing significant stimulus-related modulations based on initial permutation tests (see ■, Fig. 3A-B). For θ-band (Fig. 4A), we found elevated spectral responses when speech did not contain pitch cues (i.e., 0ST> 4ST) [$F_{1, 36} = 0.413$, $p = 0.0495$], whereas the β-band and γ_{low} -band (Fig. 4B, 4C), showed stronger oscillatory activity for clean speech (i.e., clean > noise) [β band: $F_{1, 36} = 9.73$, $p = 0.0036$; γ_{low} band: $F_{1, 36} = 5.15$, $p = 0.0294$]. Modulations in γ_{high} power oscillations (Fig. 4D) were observed

for changes in both pitch (0ST > 4ST) [$F_{1, 36} = 5.65$, $p = 0.0229$] and SNR (clean > noise) [$F_{1, 36} = 16.87$; $p = 0.0002$].

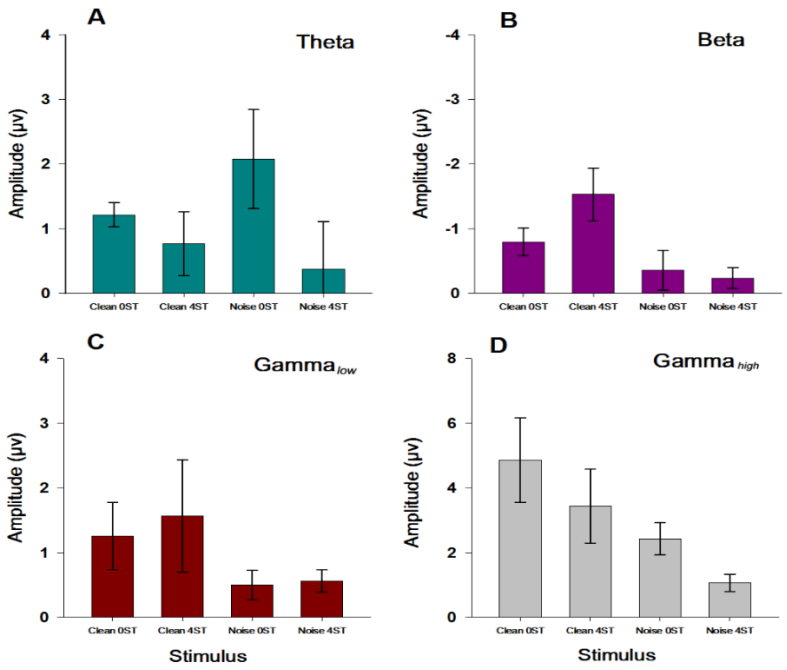

Figure 4. Band-specific mean spectral peak amplitudes across conditions. Shown here are mean amplitudes for each frequency band extracted from the temporal center of segments showing significant stimulus-related modulations (see Fig. 3). (A) θ-band spectral responses were elevated when speech did not contain pitch cues (i.e., 0ST> 4ST). (B) β-band and (C) γ_{low} -band showed stronger desynchronization for clean compared to noise-degraded speech (i.e., clean > noise). Note that negative is plotted up for this band. (D) γ_{high} power modulations were observed for changes in both pitch (0ST > 4ST) and SNR (clean > noise). error bars = ±1 s.e.m. Together, these findings demonstrate that difference in neural activity to speech between conditions is derived by acoustic features, signal quality, and the cognitive effort which causes changes in underlying low vs. high bands of oscillatory activity.

Brain-behavior relationships

Bivariate regressions between band-specific EEG amplitudes and behavioral accuracy and RTs are shown in Figure 5A and 5B, respectively. For each frequency band, we derived a singular measure of *neural* F0-benefit, computed as the change in response with and without pitch cues (e.g., $\Delta\ \beta_{4ST} - \beta_{0ST}$). This neural measure was then regressed against each listener's

behavioral F0-benefit for the accuracy and RT measures (i.e., $\Delta PC_{4ST-}PC_{0ST}$ for accuracy scores; $\Delta RT_{4ST-}RT_{4ST}$ for reaction times). Paralleling our previous work on speech perception (cf. Bidelman, 2017a; Bidelman & Walker, 2017), we reasoned that larger neural differentiation between the 0ST and 4ST would correspond to larger gains in behavioral performance (i.e., larger perceptual F0-benefit). Repeating this analysis for each band allowed us to evaluate potential mechanistic differences in how different neural rhythms map to behavior. Each matrix cell shows the regression's *t*-statistic which indicates both the magnitude and sign (i.e., negative vs. positive) of the association between variables.

These analyses revealed that γ_{low} was associated ($R^2 = 0.17$) with %-accuracy in the double vowel task when pooling clean and noise conditions. Analysis by SNR indicated that this correspondence was driven by how γ_{low} differentiated clean speech ($R^2 = 0.42$). Additional links were found between behavioral RT speeds and neural F0-benefit, particularly for low-frequency bands of the EEG. Notably, changes in θ- ($R^2 = 0.71$) and β- ($R^2 = 0.19$) oscillations predicted listeners' RTs, particularly for noise-degraded speech[1]. Collectively, these findings imply that higher frequency oscillatory rhythms (γ-band) might reflect the quality of stimulus representation and thus accuracy in identifying double-vowel mixtures. In contrast, low-frequency oscillations are associated with the speed of individuals' decisions and thus the listening effort associated with concurrent sound processing.

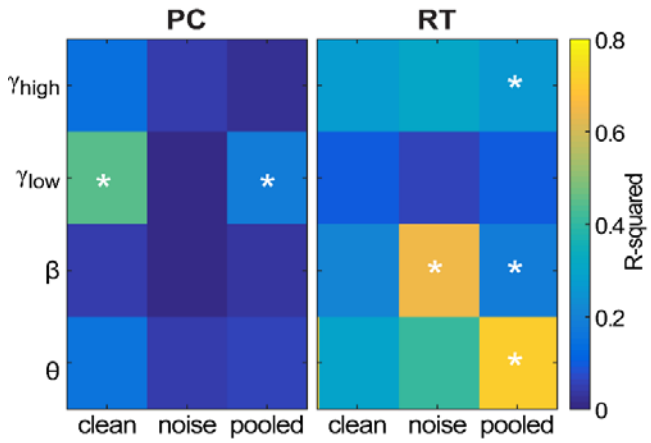

Figure 5. Brain-behavior correlations underlying double-vowel segregation. Individual cells of each matrix show the *t*-statistic for the regression indicating both the magnitude and sign of association between *neural* F0-benefit and listeners' corresponding *behavioral* F0-benefit. In both cases, larger F0-benefit reflects more successful neural/behavioral speech segregation with the addition of pitch cues (i.e., 4ST > 0ST). (A) correspondences between neural responses and identification accuracy (%); (B) correspondence with RTs. Changes in γ_{low} activity predict improved behavioral accuracy in double-vowel identification whereas the speed of listeners' decision are predicted by changes in lower oscillations (θ and β band). PC = percent correct, RT= reaction times. $*p < 0.05$, $**p \leq 0.01$, $***p \leq 0.001$.

Listeners QuickSIN scores were low (-0.73 ± 1.3 dB SNR loss), consistent with the average speech-in-noise perception abilities for normal-hearing listeners (i.e., 0 dB). QuickSIN scores were not correlated with any band-specific oscillations across SNR or pitch conditions.

DISCUSSION

The present study measured rhythmic neuroelectric brain activity as listeners rapidly identified double-vowel stimuli varying in their voice pitch (F0) and noise level (SNR). Results showed three primary findings: (i) behaviorally, listeners exploit F0 differences between vowels to segregate speech and this perceptual F0 benefit is larger for clean compared to noise degraded (+ 5dB SNR) stimuli; (ii) oscillatory power of lower θ and β frequency bands of the EEG reflects cognitive processing modulated by task demands (e.g., listening effort, memory), whereas high γ_{low} and γ_{high}-band power tracks acoustic features (e.g., envelope) and quality (i.e., noisiness) of

the speech signal (i.e., stimulus encoding); (iii) perceptual performance in segregating speech sounds is predicted by different modulatory effects in band-specific activity: low-frequency oscillations correlate with behavioral reaction times in double vowel identification whereas high-frequency oscillations are linked to accuracy. The differential changes in power across frequency bands of the EEG suggest the engagement of different brain mechanisms supporting speech segregation that vary with pitch and noise cues in auditory mixtures.

Effects of SNR and F0 cues on behavioral concurrent vowel segregation

Consistent with previous behavioral data (Arehart et al., 1997; Chintanpalli et al., 2016; Chintanpalli & Heinz, 2013a; Reinke et al., 2003), we found that listeners were better at perceptually identifying speech mixtures when vowels contained different F0s (4ST) compared to identical (0ST) F0s in both clean and noise conditions (clean > noise). This perceptual F0 benefit was larger for clean compared to noise degraded (+ 5dB SNR) stimuli. However, we extend prior studies by demonstrating that the acoustic stressor of noise limits the effectiveness of these pitch cues for segregation. Indeed, F0-benefit was weaker for double-vowel identification amidst noise compared to clean listening conditions. Similarly, smaller ΔRTs (accompanied by lower accuracy) for segregating in noise suggests that listeners experienced a time-accuracy tradeoff such that they achieved more accurate identification of speech at the expense of slower decision times (Fig. 1).

Computationally, the identification of concurrent vowels is thought to involve a two-stage process in which the auditory system first determines the number of elements present in a mixture (i.e., “1” vs. “2” sounds) and then seeks their identities (~150-200 ms). The former process (segregation) is thought to involve a comparison of the incoming periodicities of double-vowel F0s, which could be realized via autocorrelation-like mechanisms in peripheral (Bidelman

& Alain, 2015a; A. Chintanpalli, Ahlstrom, & Dubno, 2014b; Du et al., 2010; R. Meddis & Hewitt, 1992b) and/or auditory cortical neurons (Alain, Reinke, He, et al., 2005; Bidelman & Alain, 2015a; Du et al., 2010).

Indeed, neurons in primary and surrounding belt areas of auditory cortex are both sensitive to pitch and even display multi-peaked tuning with peaks occurring at harmonically-related frequencies (Bendor, Osmanski, & Wang, 2012a; Kikuchi, Horwitz, Mishkin, & Rauschecker, 2014b). Following F0-based segregation, the process of determining vowel identity could be realized via template matching mechanisms (~ 300-400 ms) in which each representation is matched against internalized memory profiles for both vowel constituents. Using a computational model of this two-stage model (i.e., autocorrelation-based segregation followed by template matching), Meddis and colleagues (R. Meddis & Hewitt, 1992b) have shown that identification of two synthesized vowels with the same F0 improves from ~40% to 70% when vowels differ in F0 from 0 to 4 ST—consistent with the F0-benefit in this study. While F0 cues are likely the primary cue for segregation in our double vowel task, conceivably, listeners might also use additional acoustic cues to parse speech such as spectral differences associated with formants (Chintanpalli & Heinz, 2013a), temporal envelope cues produced by harmonic interactions (Culling & Darwin, 1993), and spectral edges.

Cortical oscillations reveal mechanistic differences in concurrent speech segregation divisible by frequency band

It is useful to cast our behavioral data in the context of this computational framework. We found that listeners showed weaker F0-benefit when speech was presented in noise. Poor performance in the noise conditions could result either from poorer segregation at the initial front end (prior to classification) or weaker matching between the noisy vowel representations and

their speech templates. Our behavioral data do not allow us to unambiguously adjudicate these two explanations. In this regard, EEG time-frequency results help isolate different mechanistic accounts. In response to a stimulus, synchronous temporal activity is represented as multiple time courses in brain networks via EEG oscillations whose amplitude depends on the degree of neural synchrony. Different frequencies respond differently to sensory stimuli and task demands (Hanslmayr, Gross, Klimesch, & Shapiro, 2011). Stimulus rhythmic event-related activity can either increase (synchronization) or decrease (de-synchronization) as networks are either engaged or disengaged, respectively (Destexhe, Hughes, Rudolph, & Crunelli, 2007).

Presumably, the acoustic features contributing to the segregation of the speech depend on the availability of those cues to the auditory system. That is, the encoding and weighting of acoustic cues along the auditory pathway may change depending on the quality of the incoming signal. Electro-physiologically, we observed multiple, frequency-specific time courses to concurrent speech segregation with activity unfolding within different channels of the EEG dependent on both the pitch and SNR of speech. Previous M/EEG work has shown similar sequences of events in the early object negativity response (~150 ms) (Alain, Reinke, He, et al., 2005; Du et al., 2010) and early interactions of pitch and noise cues (~200 ms) (Bidelman & Yellamsetty, 2017b) followed by automatic registration of F0 differences at ~250 ms (Alain, Reinke, He, et al., 2005; Du et al., 2010).

In cases where vowel mixtures were further distorted by noise, γ_{high} power showed reduced tracking of stimulus onset/offset (cf. Ross et al., 2010). γ_{high} power was also stronger for 0ST compared to 4ST speech (i.e., mixtures which did not contain pitch cues). Higher γ activity for both clean and 0ST conditions may be due to the fact that these stimuli offer a more veridical and robust representation of the speech signal envelope; clean speech being unconcluded and

0ST vowels offering a singular harmonic structure (common F0). Under this interpretation, modulations in γ activity in our double vowel task are arguably ambiguous as they signal both cleaner signals (clean > noise) simultaneously with representations that cannot be cleanly segregated (0ST > 4ST) (cf. Fig. 3A and 3B). Relatedly, brain-behavior correlations showed that larger changes in γ activity with the addition of pitch cues were associated with *poorer* behavioral F0-benefit (Fig. 5A). Given that higher bands of oscillations are thought to reflect signal identity and the construction of perceptual objects (Bidelman, 2015a, 2017a; C. Tallon-Baudry & Bertrand, 1999b), our data suggest that the auditory brain must rely on more than object-based information for successful segregation.

In contrast to the higher γ-band modulations, we also observed distinct modulation in lower bands of the EEG that covaried with successful speech segregation. Interestingly, β band amplitudes were suppressed for easier stimulus conditions (e.g., clean 4ST; Fig. 4B), suggesting a desynchronization in this frequency range. Similarly, θ-band activity showed prominent increases (synchronization) for difficult 0ST and noise-degraded speech. β band (15-30 Hz) has been linked with the extraction of global phonetic features (Bidelman, 2015a, 2017a; Fujioka et al., 2012; Ghitza, 2011), template matching (Bidelman, 2015a), lexical semantic memory access (Shahin et al., 2009), and perceptual binding (Aissani et al., 2014; Brovelli et al., 2004; von Stein & Sarnthein, 2000). In contrast, θ-band may reflect and attention/arousal (Aftanas et al., 2001; Paus et al., 1997). Enhancements in θ-activity and suppression in β-modulations are known to correlate with the level of attention and memory load in a wide variety of tasks (Bashivan, Bidelman, & Yeasin, 2014; Bastiaansen et al., 2005; Fries, Reynolds, Rorie, & Desimone, 2001). Modulations of M/EEG amplitudes during the conscious identification of simultaneous speech occurs around ~350 to 400 ms post stimulus onset (Alain, 2007a; Alain et al., 2017; Alain,

Reinke, He, et al., 2005; Bidelman & Yellamsetty, 2017b; Du et al., 2010; Reinke et al., 2003) relating to the time course of β- and θ-band oscillatory activity observed in this study.

Thus, we suggest perceptual success in parsing multiple speech streams is driven by the degree of cognitive processing (e.g., attentional deployment, listening effort) that is determined by the availability of acoustic features and signal quality. Cleaner, less distorted speech presumably allows more successful matches between speech acoustics and internalized speech templates which would aid identification. This notion is supported by the fact that larger changes in θ responses were associated with *smaller Δ*RTs whereas larger changes in β responses were associated with *larger Δ*RTs (Fig. 5B). Given that listeners required a longer time to accurately judge double-vowels (i.e., $\Delta RT_{clean} > \Delta RT_{noise}$ time-accuracy tradeoff; Fig. 1B), the most parsimonious interpretation of our neural results are that θ-band elevates due to increased listening effort or cognitive demands of the task (e.g., conditions without F0 cues) whereas β-band decreases, reflecting easier and/or more successful memory template matching (e.g., clean speech conditions).

On the additivity vs. interactions of cues for concurrent sound segregation

Notably, while EEG measures showed a correspondence with behavior for double vowel identification, we did not observe correlations between neural measures and QuickSIN scores. However, this might be expected given differences in task complexity and the fact that the former was recorded during electrophysiological testing while the latter was not. Nevertheless, these findings corroborate our previous studies and suggest that mechanisms that exploit sequential and concurrent auditory streaming are likely independent (or at least different) from the mechanisms recruited for complex speech in noise recognition (Alain et al., 2017; Hutka, Alain, Binns, & Bidelman, 2013). For example, the QuickSIN may rely more on cognitive

(rather than perceptual) processes, such as attention, working memory, and linguistic processing, while double-vowel identification used in the present study are more perceptual-based. Future work is needed to explore the relationship (or lack thereof) between concurrent speech segregation and more generalized speech-in-noise recognition tests.

The differential changes in oscillatory θ-, β-, and γ power and F0 x SNR interaction in α- and β- bands illustrates potential differences in the brain mechanisms supporting speech segregation that are largely divisible into high- and low-frequency brain rhythms. The neural interaction of pitch and noise that are circumscribed to α- and β- bands and in the earliest time windows (~150 to 200 ms) is consistent with our previous ERP studies which revealed significant F0 x SNR interactions in concurrent vowel encoding and perception in the timeframe of the N1-P2 complex (Bidelman & Yellamsetty, 2017b). Overall, we found that different acoustic factors (SNR vs. noise) influenced the neural encoding of speech dynamically with interaction effects early but additive effects occurring later in time. Our results are partially in agreement with the additive effects on concurrent vowel perception shown by Du et al. (2011), who suggested that listeners rely on a linear summation of cues to accumulate evidence during auditory scene analysis. Indeed, our data show that high- (γ) and low- (θ) frequency responses carry independent information on speech processing later in time (>300-400 ms). However, our results further reveal that acoustic cues (here SNR and F0) can interact earlier (~100-200 ms; Fig. 3C) to impact double vowel processing. Notably, Du et al. (2011) study investigated the effects of F0 and *spatial location* on concurrent vowel perception. Given that spatial and non-spatial (cf. F0) cues are largely processed via independent information channels of the brain (i.e., dorsal and ventral pathways) (S. R. Arnott, Binns, Grady, & Alain, 2004), acoustic differences among sources might be expected to combine linearly as reported in that study (Y. Du, He, et al.,

2011). In contrast, our behavioral and electrophysical results suggest acoustic cues that affect the inherent acoustic representation of speech signals (i.e., pitch and noise) can actually interact fairly early in the time course of speech segregation and are not processed in a strictly additive manner (Bidelman & Yellamsetty, 2017b).

Directions for future work

Previous ERP studies have shown success in identifying concurrent vowels improves with training accompanied by decreased N1 and P2 latencies and enhanced P2 peak amplitude (Alain, 2007a; C. Alain et al., 2007). In future extensions of this work, it would be interesting to examine how the weighting of neural activity changes across frequency bands with perceptual learning. For example, a testable hypothesis is that neural changes in lower frequency bands might accompany top-down automatization during successful learning. We would also predict that higher frequency bands would begin showing improved signal coding with task repetition and increased familiarity with the incoming signal. Another interesting study would be to investigate multiple competing streams and how attention might modulate concurrent speech segregation (Ding & Simon, 2012; Krumbholz, Eickhoff, & Fink, 2007). Future studies are needed to test the role of band-specific mechanisms of the EEG in relation to short-term speech sound training, learning, and attentional effects on concurrent speech segregation.

CONCLUSIONS

By measuring time-frequency changes in the EEG during double vowel identification, we found band-specific differences in oscillatory spectral responses which seem to represent unique mechanisms of speech perception. Over the 200 ms stimulus duration, early envelope tracking of the stimulus duration (onset/offset) was observed in higher frequency oscillations of the γ band. This was followed by stronger desynchronization (suppression) in the mid-frequency β

oscillations around (~250 to 350 ms). Finally, differences in lower frequency θ oscillations were more pervasive and persisted across a larger extent of each trial (~400 -500 ms after stimulus onset). We infer that early portions of time-frequency activity (higher-bands) likely reflect pre-perceptual encoding of acoustic features and follow the quality of the speech signal. This capture of stimulus properties is then followed by post-perceptual cognitive operations (reflected in low EEG bands) that involve the degree of listening effort and task demands. Tentatively, we posit that successful speech segregation is governed by more accurate perceptual object construction, auditory template matching, and deceased listening effort/attentional allocation, indexed by the γ-, β-, and θ-band modulations, respectively.

Chapter 3

SUBCORTICAL PROCESSING OF CONCURRENT SPEECH MIXTURES AS REVEALED BY FREQUENCY-FOLLOWING RESPONSES

Abstract

When two voices compete, listeners can segregate and identify concurrent speech sounds using pitch (fundamental frequency, F0) and timbre (harmonic) cues. Speech perception is also hindered by the signal-to-noise ratio (SNR). How clear and acoustically-degraded concurrent speech representations in early, pre-attentive stages of the auditory system is not well understood. We recorded frequency-following responses (FFR) while listeners heard two steady-state single and double vowels- whose F0 differed by zero or four semitones (ST) presented in either clean (no noise) or noise-degraded (+5dB SNR) conditions. Listeners also performed a speeded double vowel identification task in which they were required to identify both vowels correctly. Behavioral results showed that speech identification accuracy increased with F0 differences between vowels, and this perceptual F0 benefit was larger for clean compared to noise degraded (+ 5dB SNR) stimuli. Neurophysiological data demonstrated more robust FFR F0 amplitudes for single compared to double vowels and considerably weaker responses in noise. F0 amplitudes showed speech-on-speech masking effects along with a non-linear constructive interference at 0ST, and suppression effects at 4ST. Correlations showed that FFR F0 amplitudes failed to predict listeners identification accuracy. In contrast, FFR F1 amplitudes were associated with faster reaction times, although this correlation was limited to the noise condition. The limited number of brain-behavior associations suggests subcortical activity mainly reflects exogenous processing rather than perceptual correlates of concurrent speech perception.

Collectively, our results demonstrate that FFRs reflect the pre-attentive mechanisms of concurrent stimulus interactions that weakly predict the success of identifying simultaneous speech.

Keywords: FFR; double-vowel identification; speech-in-noise perception

INTRODUCTION

A fundamental phenomenon in human hearing is the ability to parse co-occurring auditory objects (e.g., different voices) to extract the intended message of a target signal. Psychophysical and neurophysiological studies have shown that listeners can use multiple cues to distinguish simultaneous sounds. The segregation of a complex auditory mixture is thought to involve a multistage hierarchy of processing, whereby initial pre-attentive processes that partition the sound waveform into distinct acoustic features (e.g., pitch, harmonicity) are followed by later, post-perceptual Gestalt principles (Koffka, 1935) (e.g., grouping by physical similarity, temporal proximity, good continuity (Bregman, 1990b) and phonetic template matching (Alain, Reinke, He, et al., 2005; R. Meddis & Hewitt, 1992b). Psychophysical research from the past several decades confirms that human listeners exploit fundamental frequency (F0) (i.e., pitch) differences to segregate concurrent speech (Arehart et al., 1997; Assmann & Summerfield, 1989b, 1990b, 1994b; Chintanpalli et al., 2016; de Cheveigné et al., 1997). For example, when two steady state synthetic vowels are presented simultaneously to the same ear, listeners' identification accuracy increases significantly when a difference of four semitone (ST) is introduced between their F0s (Assmann & Summerfield, 1989a, 1990a, 1994a; Culling, 1990; McKeown, 1992; Scheffers, 1983; Zwicker, 1984). This improvement is referred to as the "F0-benefit" (Arehart et al., 1997; Bidelman & Yellamsetty, 2017b; Chintanpalli, Ahlstrom, &

Dubno, 2014a; Chintanpalli et al., 2016; Chintanpalli & Heinz, 2013a; Yellamsetty & Bidelman, 2018a).

To understand the time course of neural processing underlying concurrent speech segregation most investigations have quantified how various acoustic cues including harmonics, spatial location, and onset asynchrony affect perceptual segregation (Claude Alain, 2007b; Carlyon, 2004). However, the overwhelming majority of neuroimaging studies have been concerned with the *cortical* representations/correlates of concurrent speech perception (Alain, Reinke, McDonald, et al., 2005; Bidelman, 2015a; Bidelman & Yellamsetty, 2017b; Dyson & Alain, 2004; Yellamsetty & Bidelman, 2018a). In contrast, the *subcortical* neural underpinnings of segregation have been studied only in animals. Studies that directly examined the representation of F0's of concurrent complex tones in auditory nerve (AN) and cochlear nucleus (CN) showed the temporal discharge pattern and the spatial distribution of AN and CN fibers contain sufficient information to identify both F0s (Jane & Young, 2000; Keilson et al., 1997; Palmer, 1990b; Alan R Palmer & Winter, 1992; Sinex, 2008; Tan & Carney, 2005). The same is observed for double vowel speech stimuli (Keilson et al., 1997; Palmer, 1990b; Alan R Palmer & Winter, 1992). In addition, AN single-unit population studies have shown neural phase-locking is a primary basis for encoding the tonal features (e.g., F0) of vowels (Reale & Geisler, 1980; Tan & Carney, 2005) and that different sets of neurons are involved in encoding the first and second formants of speech (Miller, Schilling, Franck, & Young, 1997). Whereas at the level of the inferior colliculus (IC), responses are tuned to low-frequency amplitude fluctuations (Bidelman & Alain, 2015a; Sinex, 2008; Sinex, Henderson, et al., 2002; Sinex et al., 2005; Sinex, Sabes, et al., 2002b), providing a robust neural code for both F0 periodicity and the spectral peaks (i.e., formants) that listeners use to separate and identify vowels (Carney et al.,

2015; Henry et al., 2017). These temporal discharge patterns are closely related to the autocorrelation model of pitch extraction (Ray Meddis & Hewitt, 1992c) that accounts for the encoding of single and multiple F0s at the level of AN (Cariani & Delgutte, 1996; Cedolin & Delgutte, 2005; Ray Meddis & Hewitt, 1992c). It appears that stimulus periodicity (F0) are coded very early in the auditory system and remain largely untransformed in the phase-locked activity of the rostral brainstem (Bidelman, 2015a). Thus, evoked potentials, which measure phase-locked brainstem activity, could offer a window into how subcortical regions of the *human* brain encode concurrent sounds, including those based on F0-segregation (i.e., double-vowel mixtures).

In the present study, we used the scalp-recorded human frequency-following response (FFR), which reflects sustained phase-locked activity dominantly from the rostral brainstem (Bidelman, 2018; Glaser, Suter, Dasheiff, & Goldberg, 1976; Marsh, Brown, & Smith, 1974; Smith, Marsh, & Brown, 1975; Worden & Marsh, 1968). FFRs can reproduce frequencies of periodic acoustic stimuli below approximately 1500 Hz (Bidelman & Powers, in press-a; Gardi, Merzenich, & McKean, 1979; Stillman, Crow, & Moushegian, 1978). FFRs code important properties of speech stimuli such as voice F0 (Bidelman, Gandour, & Krishnan, 2011; Krishnan, Bidelman, & Gandour, 2010) and several lower speech harmonics (formants) (Bidelman, 2015b; B Chandrasekaran & Kraus, 2010; Krishnan, 1999, 2002b; Krishnan & Agrawal, 2010). This allows us to estimate how salient properties of speech spectra (e.g., F0s or formants of concurrent vowels) are transcribed by the *human* auditory nervous system at early, pre-attentive stages of the processing hierarchy.

In addition, FFRs have provided critical insight toward understanding the neurobiological encoding of degraded speech from a subcortical perspective (Anderson, Skoe, Chandrasekaran,

Zecker, & Kraus, 2010a; Bidelman, 2017b; Bidelman & Krishnan, 2010a; A Parbery-Clark, Skoe, & Kraus, 2009; Song, Skoe, Banai, & Kraus, 2011). Speech perception in noise is related to the subcortical encoding of F0 and timbre (Bidelman, 2016a; Bidelman & Krishnan, 2010a; Song et al., 2011) as well as the effectiveness of the nervous system to extract regularities in speech sounds related to vocal pitch (Chandrasekaran, Hornickel, Skoe, Nicol, & Kraus, 2009; Xie, Reetzke, & Chandrasekaran, 2017). Resilience of the FFR at F0 (but not its higher harmonics or onset) in the presence of noise has been noted by a number of investigators (Bidelman & Krishnan, 2010a; Li & Jeng, 2011; Prévost, Laroche, Marcoux, & Dajani, 2013; Russo, Nicol, Musacchia, & Kraus, 2004) and suggests that neural synchronization at the fundamental F0 periodicity is relatively robust to acoustic interference (for review, see (Bidelman, 2017b))—at least for *single* speech tokens presented in isolation.

Given its remarkable spectro-temporal fidelity, we reasoned that neural correlates relevant to double vowel identification may be substantiated in nascent signal processing along the acoustic pathway, even earlier than documented in cerebral cortex (Alain et al., 2017; Alain, Reinke, He, et al., 2005; Bidelman & Yellamsetty, 2017b; Yellamsetty & Bidelman, 2018a). We aimed to test this hypothesis by analyzing the spectral response patterns of the single and double vowel FFRs when speech sounds did and did not contain distinct F0 cues (0ST vs. 4ST). Additionally, we examined concurrent vowel processing in different levels of noise interference (quiet vs. +5 dB SNR) to evaluate how the neural encoding of spectro-temporal cues might interact with noise at a subcortical level. Despite ample FFR studies using isolated speech sounds (e.g., vowels, stop consonants) (Anderson & Kraus, 2010; Bidelman & Krishnan, 2010a; Hornickel, Skoe, Nicol, Zecker, & Kraus, 2009; Krishnan, 2002a; A Parbery-Clark, Skoe, &

Kraus, 2009), to our knowledge, this is the first to examine brainstem encoding of speech mixtures in the human auditory system.

Here, we sought to (1) determine how concurrent vowels are encoded at pre-attentive, subcortical levels of the auditory nervous system; (2) characterize the effects of noise on the neural encoding of voice pitch and timbre (i.e., formant) cues in concurrent speech; and (3) establish the relation between (pre-attentive) brainstem neural activity and behavioral concurrent vowel identification in quiet and degraded listening conditions. To this end, we recorded neuroelectric responses as listeners passively heard double-vowel pairs and single vowel stimuli. Stimulus manipulations were designed to promote or deny successful identification (i.e., changes in F0 separation of vowels; with/without noise masking). We expected the spectral components of FFRs to reflect the encoding of non-linear interactions between the two concurrent vowels, such that responses would differ with and without pitch cues in a constructive and suppressive manner. Additionally, we hypothesized FFRs would show reduced amplitudes with noise and correlate with behavioral identification scores, offering an objective, subcortical correlates of concurrent speech perception.

EXPERIMENTAL PROCEDURES

Participants

Sixteen young adults (age $M \pm SD$: 24 ± 2.25 years; 10 females, 6 males) participated in the experiment. All the participants had obtained a similar level of formal education (18.18± 2.16 years), were right handed (>43.2% laterality) (Oldfield, 1971), had normal hearing thresholds (i.e., ≤ 25 dB HL) at octave frequencies between 250 and 8000 Hz, and reported no history of neuropsychiatric disorders. Each gave written informed consent in compliance with a protocol approved by the University of Memphis Institutional Review Board.

Stimulus and Behavioral task

Double vowel stimuli

Speech stimuli for FFR recordings were modeled after previous studies on concurrent double-vowel segregation (Alain, 2007a; Assmann & Summerfield, 1989a, 1990a; Bidelman & Yellamsetty, 2017b; Yellamsetty & Bidelman, 2018a). Synthetic, steady-state vowel tokens (/a/ and /ɛ/) are created using a Klatt synthesizer (Klatt, 1980) implemented in MATLAB® 2014 (The MathWorks, Inc.). Each token was 200 ms in duration including 10-ms $\cos^2$ onset/offset ramping. F0 was either 150 or 190 Hz and formant frequencies (F1, F2) were 766 Hz, 1299 Hz and 542 Hz, 1780 Hz for /a/ and /ɛ/, respectively. These F0s were selected since they are above the frequencies of observable FFRs in cortex (Bidelman, 2018; Brugge et al., 2009), and thus ensured responses would be of brainstem origin (Bidelman, 2018). Double-vowel stimuli were then created by combining single-vowels in pairs. Each vowel pair had either identical (0ST) or different F0s (4ST). That is, one vowel's F0 was set at 150 Hz while the other vowel had an F0 of 150 or 190 Hz so as to produce double-vowels with an F0 separation of either 0 or 4 semitones (STs), resulting in two double-vowel pairs (1 pair x 2 F0 combinations).

Both single and double-vowels were presented in clean and noise conditions (separate blocks), in which stimuli were delivered concurrent with a backdrop of multi-talker noise babble (+5 dB SNR) (Bidelman & Howell, 2016; Nilsson et al., 1994). SNR was manipulated by changing the level of the masker rather than the signal to ensure that SNR was not positively correlated with overall sound level (Bidelman & Howell, 2016; Binder et al., 2004). Babble was presented continuously to avoid it time-locking with stimulus presentation. We chose continuous babble over other forms of acoustic inference (e.g., white noise) because it more closely mimics

real-world listening situations and tends to have a larger effect on the auditory ERPs (Kozou et al., 2005).

Behavioral double-vowel identification task.

Participants were presented with double-vowel combination of synthetic steady-state vowel tokens (/a/, /ɛ/, and /u/) as in our previous studies (Bidelman & Yellamsetty, 2017b; A. Yellamsetty & Bidelman, 2018b). Double-vowels were presented in separate blocks of clean and noise (+5 dB SNR) conditions. Listeners were asked to identify both vowels as quickly and accurately as possible on the keyboard. Feedback was not provided.

Prior to the experiment proper, we required that participants be able to identify single vowels in a practice run with >90% accuracy (e.g., C. Alain et al., 2007). This ensured their task performance would be mediated by *concurrent* sound segregation skills rather than isolated identification, *per se.*

FFR data recording and preprocessing

For the FFR recordings, participants reclined comfortably in an IAC electro-acoustically shield booth. Participants were instructed to relax and refrain from extraneous body movements while they watched a muted subtitled movie (i.e., passive listening task). EEGs were recorded differentially between Ag/AgCl disk electrodes placed on the scalp at the high forehead (~Fpz) referenced to link mastoids A1/A2) and forehead electrode as ground. Interelectrode impedances were maintained <2 kΩ. Stimulus presentation was controlled by MATLAB routed to a TDT RP2 interface (Tucker-Davis Technologies). Speech stimuli were delivered binaurally using fixed (rarefaction) polarity at an intensity of 81 dB SPL through shielded ER-2 insert earphones (Etymotic Research). Control runs confirmed no artifacts in the FFR response waveforms. The order of single and double vowel stimuli was randomized within and across participants; clean

and noise conditions were run in separate blocks. The inter-stimulus interval was 50 ms. In total, there were 2000 trials for each of the individual stimulus conditions.

Neural activity was digitized using a sampling rate of 10 kHz (SynAmps RT amplifiers; Compumedics Neuroscan). EEGs were then epoched (0-250 ms) and averaged in the time domain to derive FFRs for each condition. Sweeps exceeding ±50 μV were rejected as artifacts prior to averaging. FFRs were then bandpass filtered (100 to 3000 Hz) for response visualization and quantification. The entire experimental protocol including behavioral and electrophysiological testing took 2.5 hrs to complete.

FFR analysis

Fast Fourier transforms (FFTs) were computed from the response time-waveforms (0 to 250 ms) using Brainstorm (V.3.4) (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011a). Brainstorm expresses FFT amplitudes as power with a scaling factor of units^2 /Hz * 10^{-13}; subsequent measures reflect this scaling. From each FFR spectrum, we measured the F0, harmonics, and F1-formant frequency amplitudes to quantify "pitch" and "timbre" coding for each condition. We estimated the magnitude of the response at F0 and harmonics of the single and double vowels by manually picking the maximum spectral energy within 10 Hz wide bins surrounding the F0 and five harmonics. F1 magnitude was taken as the average spectral energy (on a linear scale) in the frequency ranges between 392- 692 Hz for /ɛ/$_{150Hz}$ (0ST), 352- 732 Hz for /ɛ/$_{190Hz}$ (4ST) and 616-916 Hz for /a/$_{150Hz}$ vowels. These ranges were determined based on the expected F0/F1 frequencies from the input stimulus. Stimulus-related changes in F0 and F1-formant magnitudes provide an index of how concurrent stimuli and noise interference degrade the brainstem representation of pitch and timbre cues in speech.

Behavioral data analysis

Identification accuracy and the "F0 benefit"

Behavioral identification was analyzed as the percent of trials where *both* vowel sounds were correctly identified. Percent correct scores were arcsine transformed to improve homogeneity of variance assumptions necessary for parametric statistics (Studebaker, 1985). Increasing the F0 between two vowels provides a pitch cue which leads to an improvement in accuracy identifying concurrent vowels (Assmann & Summerfield, 1990b; Chintanpalli & Heinz, 2013a; R. Meddis & Hewitt, 1992b).

Reaction time (RTs)

For a given double-vowel condition, behavioral speech labeling speeds [i.e., reaction times (RTs)] were computed separately for each participant as the median response latency across trials. RTs were taken as the time lapse between the onset of the stimulus presentation and listeners' identification of both vowel sounds. RTs shorter than 250 ms or exceeding 6000 ms were discarded as implausibly fast responses and lapses of attention, respectively (e.g., Bidelman & Yellamsetty, 2017b; A. Yellamsetty & Bidelman, 2018b).

Statistical analysis

Unless otherwise noted, two-way, mixed-model ANOVAs were conducted on all dependent variables (GLIMMIX Procedure, SAS® 9.4, SAS Institute, Inc.). Stimulus SNR (2 levels; clean, +5 dB noise) and semitones (2 levels; 0ST, 4ST) functioned as fixed effects; subjects served as a random factor. Tukey-Kramer multiple comparisons-controlled Type I error inflation. An *a priori* significance level was set at $\alpha=0.05$. To examine the degree to which neural responses predicted behavioral speech perception, we performed weighted least square

regression between listeners' FFRs amplitudes and perceptual identification accuracy and in the double-vowel task. Robust bisquare fitting was achieved using "fitlm" in MATLAB.

RESULTS

Behavioral data

Behavioral speech identification accuracy and RTs for double-vowel identification are shown in Figure 6. Listeners obtained near-ceiling performance (97.9 ± 1.4%) when identifying single vowels. In contrast, double-vowel identification was considerably more challenging; listeners' accuracy ranged from ~45 – 70% depending on the presence of noise and pitch cues (Fig. 6A). An ANOVA conducted on behavioral accuracy confirmed a significant SNR x F0 interaction [$F_{1, 45} = 5.65$, $p = 0.0218$], indicating that successful double-vowel identification depended on both noise and F0 pitch cues. Performance increased ~30% across the board with greater F0 separations (i.e., 4ST > 0ST). F0-benefit was larger for clean relative to +5 dB SNR speech [$t_{15} = -6.49$, $p < 0.0001$ (one-tailed)], suggesting listeners were more successful using pitch cues when segregating clean compared to speech in noise.

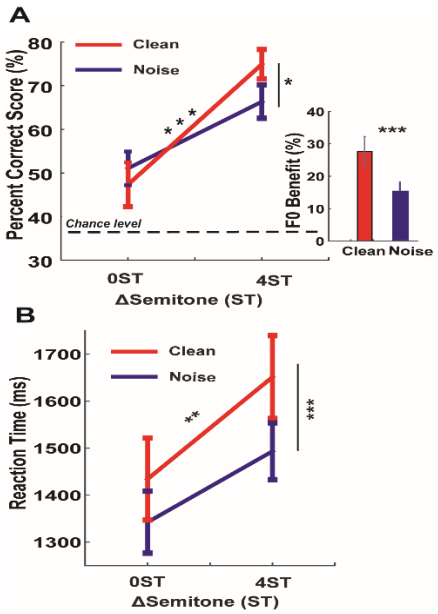

Figure 6. Behavioral responses for double-vowel stimuli. (A) Accuracy for identifying both tokens of a two-vowel mixture. Performance is poorer when concurrent speech sounds contain the same F0 (0ST) and improve ~30% when vowels contain differing F0s (4ST). (*Inset*) Behavioral F0-benefit, defined as the improvement in %-accuracy from 0ST to 4ST, indexes the benefit of pitch cues to speech identification. F0-benefit is stronger for clean vs. noisy (+5 dB SNR) speech indicating that listeners are poorer at exploiting pitch cues when segregating acoustically-degraded signals. (B) Speed (i.e., RTs) for double-vowel identification. Listeners are marginally faster at identifying speech in noise. Faster RTs at the expense of poorer accuracy (panel A) suggests a time-accuracy tradeoff in double-vowel identification. error bars = ±1 s.e.m. *p<0.05, **p ≤0.01, ***p≤0.0001.

Analysis of RTs revealed a significant effect of SNR [$F_{1, 45}$ = 16.23, p =0.0002] and ST [$F_{1, 45}$ = 7.48, p = 0.0089]; listeners tended to be slower identifying clean compared to noisy speech (Fig. 6B). The slowing of RTs coupled with better %-identification for clean compared to noise-degraded speech indicates a time-accuracy tradeoff in perception (Bidelman & Yellamsetty, 2017b; A. Yellamsetty & Bidelman, 2018b).

FFR responses to single and double vowels

Grand average FFR waveforms and spectra are shown for each vowel type (single, double vowels), SNRs (clean, noise), and semitones (0 ST, 4 ST) conditions in Figs. 7A and B.

FFRs showed phase-locked energy corresponding to the periodicities of the acoustic speech signals. Comparisons across conditions suggested more robust encoding of single and double vowels in the 0ST condition. Responses were weaker for conditions with 4ST and noise. Response spectra contained energy at the F0 and the integer-related multiples up to the upper limit of the brainstem phase locking (~1100 Hz) (Bidelman & Powers, in press-b; Liang-Fa Liu, Palmer, & Wallace, 2006). Also apparent is an apparent boost in response energy near the F1, demonstrating greater neural synchrony to formant-related harmonics. This effect is reminiscent of the formant capture phenomenon observed in peripheral auditory nerve responses (Miller et al., 1997), which acts to enhance temporal representations of spectral shape (Eric D Young & Sachs, 1979b).

Quantification of FFR F0 (pitch) and F1 (timbre) coding of single and double vowels at 0 $ST_{(/a+\varepsilon/150)}$ and 4 $ST_{(/a/150,\ /\varepsilon/190)}$ are shown in Fig. 7C. We first evaluated the effects of having multiple vs. single vowels and the effects of noise on FFR responses. A two-way mixed model ANOVA with stimulus type (2 levels: single and double vowel) and SNR (2 levels: clean and +5 dB SNR) as fixed factors (subjects= random effect) revealed that F0 amplitudes of the single-vowels were more robust than in double-vowels (single>double) [$F_{1,141}$=16.02, p<0.0001]. With noise, double-vowels showed greater reduction in F1 amplitudes than the single-vowels [$F_{1,141}$ =89.11, p<0.0001]. Responses were also stronger for double-vowels without pitch cues (i.e., 0 ST > 4 ST) revealing a super-additive effect at F0 (i.e., common F0 between vowels sum constructively in the FFR).

Next, we evaluated the impact of noise and pitch cues on *double-vowel* FFRs. Both additive and masking effects were observed at 4 ST. An ANOVA conducted on F0 amplitudes

showed significant effects of SNR [$F_{1,77}$ =31.66; p<0.0001] and ST [$F_{1,77}$ =5.67; p=0.0198] with an interaction of SNR × ST [$F_{1,77}$ =10.39; p=0.0019].

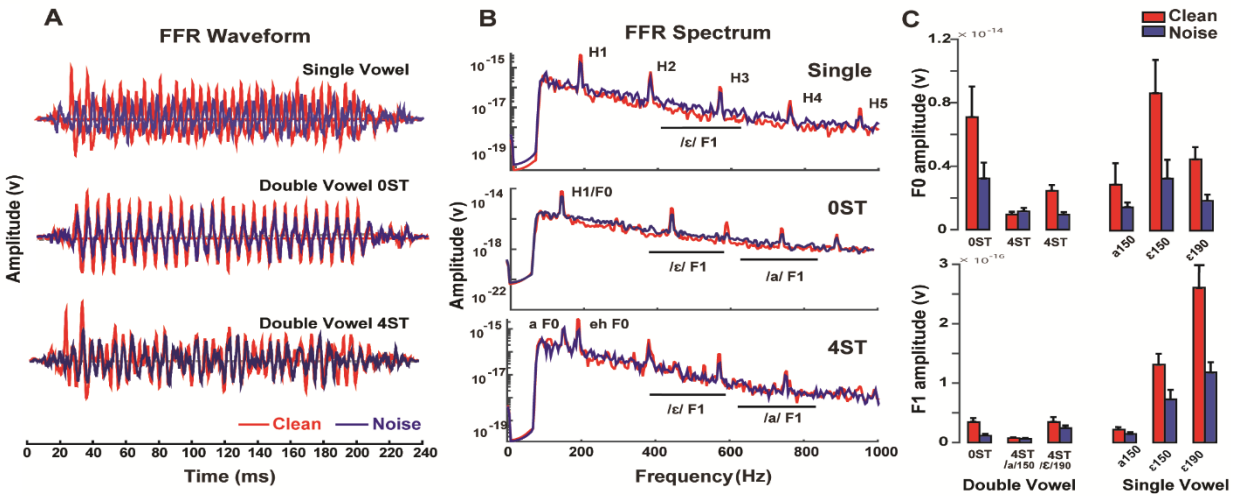

Figure 7. Brainstem FFR to double vowel mixtures. (A) FFR waveforms (B) spectra. Neural responses reveal energy at the voice fundamental (F0) and integer-related harmonics (H1-H5). F1, first formant range. (C) Brainstem encoding of the pitch (F0) and timbre (F1) as a function of the vowel count (i.e., single vs. double) and SNR. FFRs are more robust for (i) single than double vowels (single > double)—indicative of suppression (speech-on-speech masking), and (ii) at 0ST vs. 4ST (0ST > 4ST). Responses also deteriorate with noise (i.e., lower SNR). error bars = ± 1 s.e.m.

In contrast, for the neural encoding of F1, we found significant effects of ST [$F_{1,77}$ =138.15; p<0.0001] and SNR [$F_{1,77}$ =15.09; p=0.0002] but no interaction [$F_{1,77}$ =1.42; p=0.236]. Noise-related changes at F0 were greater when there were no pitch cues (0ST > 4 ST), whereas changes in F1 were greater when pitch cues were present (4ST > 0ST).

To quantify speech-on-speech masking effects in the FFR from having two vs. one vowel we assessed differences between responses to actual double vowel mixtures (i.e., $0ST_{(/a+ɛ/150)}$ and 4 $ST_{(/a/150+/ɛ/190)}$) and those evoked by the summed responses to the individual vowel constituents [e.g., $FFR_{/a-ɛ/} \geq FFR_{/a/+/ɛ/}$] (Fig. 8). The rationale of this analysis is that when multiple speech components fall within the same auditory filter band (e.g., 0ST condition), this can result in

speech-on-speech masking. The amplitude difference reflects the degree of speech-on-speech masking or mutual suppression from having two vowels in double vowel pairs. Speech-on-speech masking effects were observed in both clean (t_{15}= 2.81; p= 0.0132) and noise (t_{15}= 3.46, p= 0.0035) conditions. Suppression-like effects were observed in 4ST (in addition to speech-on-speech masking) resulting in further reduction in amplitude in both clean (t_{15}= -3.97; p= 0.001) and noise (t_{15}= -2.36; p= 0.0325). These effects were not observed at F1 (ps >>0.05). The effect of speech-to-noise (i.e., FFR amplitudes of clean vs. noise) was greater than the speech-on-speech masking (single vs. double) at F0 and F1 [$F_{1,140}$ =30.85; p<0.0001; $F_{1,140}$ =275.31; p<0.0001]. These differences indicate that FFRs to concurrent speech stimuli were systematically different than their single vowel counterparts, which also varied as a function of frequency component (i.e., F0, F1) and noise SNR.

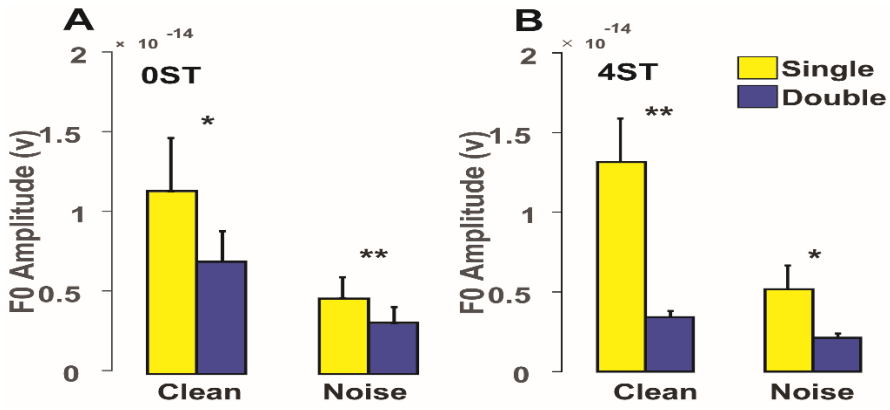

Figure 8. Additive noise vs. speech-on-speech masking effects at 0ST (A) and 4ST (B). Neural encoding of F0 for single vs. double vowels for the (A) 0 ST and (B) 4 ST mixtures. At 0 ST, (within channel) responses reflect constructive interference (additive effect) due to the same F0s and speech-on-speech masking between vowels. At 4 ST (across channel), additional suppression is observed along with the speech-on-speech masking resulting in further reduction in amplitude in both clean and noise conditions. The masking of babble noise on speech (clean vs. noise) was greater than the speech-on-speech masking (i.e., double vs. single vowel) at both 0ST and 4ST. error bars = ±1 s.e.m. *p<0.05, **p ≤0.01.

Brain-behavior relationships

Regression analyses. Pooling across ST conditions, linear regressions between FFR F0 amplitudes and behavioral accuracy (%) are shown in Figure 9A. Correlations between FFR F1 and behavioral RTs are shown in Figure 9B. We chose these analyses based on previous literature showing robust correlations between (i) FFR F0 and accuracy (Anderson, Parbery-Clark, White-Schwoch, & Kraus, 2012; Anderson et al., 2010a; Bidelman & Krishnan, 2010a; Coffey, Chepesiuk, Herholz, Baillet, & Zatorre, 2017; Du, Kong, Wang, Wu, & Li, 2011) and (ii) FFR F1 and RTs (Bidelman, Villafuerte, Moreno, & Alain, 2014; Bidelman, Weiss, Moreno, & Alain, 2014) in various speech perception tasks. These analyses revealed that F1 amplitude was associated with RTs in the noise condition (R^2 = 0.10, p=0.0277). No other correlations reached significance.

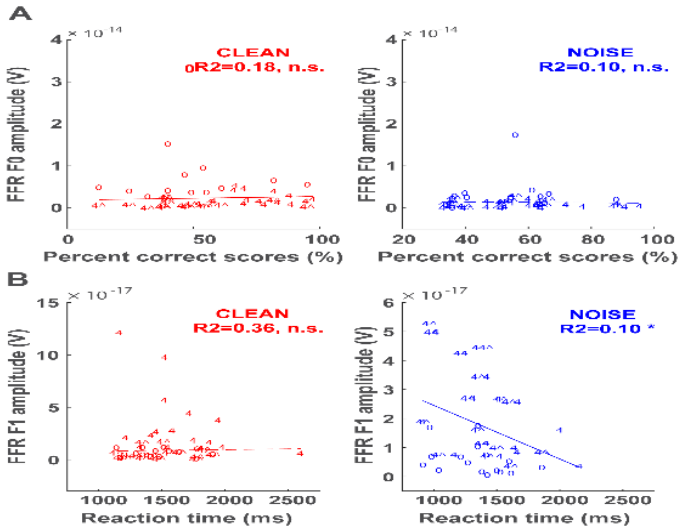

Figure 9. Brain-behavior correlations underlying double-vowel perception. Scatter plots and linear regression functions showing the relationship between (A) FFR F0 amplitudes and behavioral accuracy and (B) FFR F1 amplitudes and behavioral RTs for clean and noise-degraded speech. Data points are labeled according to each condition ('0'=0ST; '4^' = 4ST @ 150 Hz; '4' = 4ST @ 190Hz). *p<0.05, n.s. – non-significant.

Vowel dominance analysis. As an alternate approach to investigate possible relations between subcortical coding and behavioral identification of concurrent vowel mixtures we assessed whether listeners' tendency to report one or another vowel in a speech mixture depended on their FFR. We reasoned that the relative neural dominance of each (single) vowel in their double-vowel response might drive which vowel was more perceptually dominant. To quantify the relative weighting of each vowel in the FFR we carried out response-to-response Pearson's correlations between each listener's (individual) single-vowel FFR spectra (FFR_a, FFR_ε) and their double-vowel response spectrum ($FFR_{a+\varepsilon}$). The analysis is carried out at 4 ST clean condition for each participant. This analysis thus assessed the degree to which listeners' FFR to a double-vowel mixture more closely resembled a response to either /a/ or /ε/.

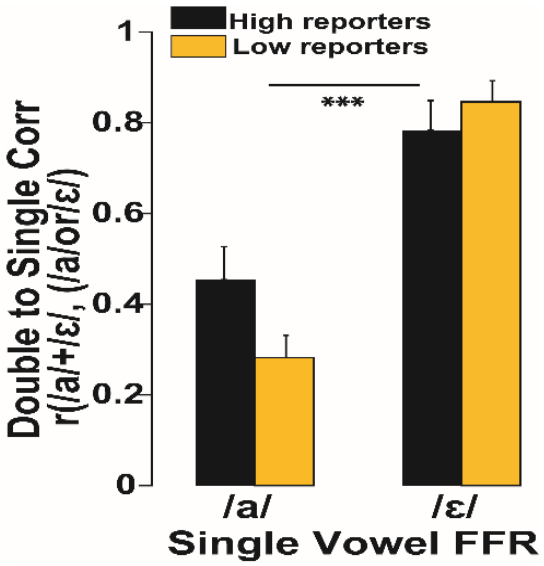

Figure 10. FFRs are modulated by stimulus salience rather than perceptual dominancy. Response-to-response Pearson's correlations between each listeners' (individual) single-vowel FFR spectra (FFR_a, FFR_ε) and their double-vowel response spectrum ($FFR_{a+\varepsilon}$). Shown here are the clean, 4 ST responses. The group split is based on the median highest and lowest 50% of listeners reporting /a/ (or /ε /) in the behavioral identification. Regardless of listeners' perceptual bias, FFRs showed better correspondence to the /ε/ vowel stimulus than /a/. error bars = ±1 s.e.m. ***$p \leq 0.0001$.

Listeners were then median split based on the counts of the highest and lowest 50% reporting /a/ in the behavioral identification task. Similarly, we determined the highest and lowest /ɛ/ reporters. We then conducted two ANOVA on response-to-response correlations with factors group vs. vowel. Figure 10 shows the response-to-response correlations with the sample split by their behavioral bias. Comparing the relative strength of response-to-response correlations, double-vowel FFRs showed better correspondence to /ɛ/ than /a/ overall. We found a vowel x group interaction ($F_{1,14}$ =4.81; p=0.0457). Even though there was a significant difference in reporting /a/ vs. /ɛ/ vowels ($F_{1,14}$ =42.89; p<0.0001) in /a+ɛ/ mixture, FFRs more closely resembled the /ɛ/ response, counter to our hypothesis.

DISCUSSION

The present study measured subcortical FFRs to double vowel stimuli that varied in their voice pitch (F0 separation) and noise level (SNR). Our results showed three primary findings: (1) behaviorally, listeners exploit F0-differences between vowels to identify speech, and the perceptual F0 benefits degrade with noise; (2) FFRs amplitudes for dual speech stimuli are altered in a systematic manner from their single vowel counterparts as a function of frequency components (i.e., F0, F1) and noise (SNR); (3) FFRs predict perceptual speed but not the accuracy of double vowel identification, but only in noisy listening conditions.

Effects of SNR and F0 cues on behavioral concurrent vowel identification

The effects of F0 on concurrent vowel identification were comparable and consistent with previous data (Arehart et al., 1997; Bidelman & Yellamsetty, 2017b; Chintanpalli et al., 2016; Chintanpalli & Heinz, 2013a; Reinke et al., 2003; Yellamsetty & Bidelman, 2018a); listeners were better at perceptually identifying speech mixtures when vowels contained pitch cues. However, we also showed that this perceptual F0-benefit was larger for the clean than the noise

degraded (+ 5 dB SNR) conditions. Additive noise tends to obscure the salient audible cues that are normally exploited by listeners for comprehension of speech (Bidelman, 2016b; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Swaminathan & Heinz, 2012). Our results indeed showed F0-benefit was weaker for double vowel identification in noise compared to clean listening condition (clean>noise). The identification of both the vowels improved from ~40% to 70% from 0 to 4 ST (Fig.6A), consistent with previous studies (Meddis & Hewitt, 1992a). We also found that RTs for identifying both vowels were faster in noise but these speeds were accompanied by lower accuracy. The longer duration RTs and more accurate identification in clean listening conditions suggests listeners experienced a time-accuracy-tradeoff (i.e., more accurate identification at the expense of slower decision times) during double vowel perception (Bidelman & Yellamsetty, 2017b; Yellamsetty & Bidelman, 2018a).

Subcortical encoding of single vs. double vowels

FFRs to single vowels showed more robust encoding than double vowels. For concurrent stimuli that do not have pitch cues (i.e., 0ST conditions with common F0's) the information for identifying the vowels is carried in the FFR only by the F1s. The improvement in the identification with pitch cues is presumably due to the more distinct timbral representations between vowels with the additional F0 separation. The pattern of nonlinear harmonic interactions in double vowels with the same two F0's (0 ST) would differ from 4 ST. At 0 ST, harmonics of both vowels fall within the same auditory filter channel and thus can add in a constructive manner. However, these within channel interactions also produce simultaneous speech-on-speech masking that results in reduced F0 amplitude for double compared to single vowels (Fig. 8A). At 4 ST, vowel harmonics fall in different auditory filters resulting in energy being spread between channels leading to a further reduction in amplitudes (Fig. 8B). Mechanistically, this

additional amplitude reduction could reflect the nonlinear phenomena of suppression (Ruggero, Robles, & Rich, 1992; Sachs & Kiang, 1968). Indeed, the ratio of our F0's at 4 ST is 1.26 (190 Hz/150 Hz), a frequency separation known to produce optimal suppression effects (Houtgast, 1974; Shannon, 1976). The spread of synchrony within/across channels most likely reflects nonlinear signal processing that helps in the identification of both vowels. In addition to the non-linearity at F0, the acoustic structure of vowels and formant-based synchrony (Delgutte & Kiang, 1984; Palmer, 1990a; Sinex & Geisler, 1983; Young & Sachs, 1979a) to harmonics near the formant (Carney et al., 2015; Miller et al., 1997; Tan & Carney, 2005; Young & Sachs, 1979a) can further sharpen the temporal representation of spectral shape in neural responses (Young & Sachs, 1979a). This may be one reason why F1-based cues are somewhat more resilient to noise than their F0 counterparts (see also [60]).

Noise effects Noise tends to obscure amplitude modulations in speech that are essential for its comprehension (Bidelman, 2016b; Shannon et al., 1995; Swaminathan & Heinz, 2012). In contrast, in cases of speech-on-speech masking, listeners can better utilize spectral dips for perception, resulting in less effective masking than continuous noise (Peters, Moore, & Baer, 1998; Shetty, 2016). Noise-related changes in FFRs were evident in both the time and frequency domain (Fig. 7A-B). Both F0 and higher spectral components (e.g., formant-related harmonics) are systematically degraded with noise, paralleling their deterioration behaviorally (Liu & Kewley-Port, 2004).

Subcortical correlates of double vowel perception

Our study showed only weak links between subcortical neural activity and behavioral percepts in the double vowel paradigm. FFRs failed to predict listeners' identification accuracy. In contrast, FFR F1 amplitudes were associated with faster RT speeds, although this correlation

was limited to the noise condition (Fig. 9B). These results replicate previous FFR studies which have shown correlations between F1 coding and behavioral RTs for speech perception (Bidelman, Villafuerte, et al., 2014; Bidelman, Weiss, et al., 2014). Yet, the F0 results contrast a large literature that has shown robust correlations between FFR F0 and degraded speech perception (Anderson et al., 2012; Samira Anderson, Skoe, Chandrasekaran, Zecker, & Kraus, 2010b; Coffey et al., 2017; Du, Kong, et al., 2011; A Parbery-Clark, Skoe, & Kraus, 2009). However, one important difference between this and previous work is that all speech-FFR studies to date have used single, isolated speech tokens (e.g., vowels, CVs) rather than the more complex double-vowel mixtures used here. Additionally, our stimuli were designed to have relatively high F0s (150 Hz), compared to other FFR studies where tokens predominantly had voice pitches of ~100 Hz. This is an important distinction as recent studies have shown that FFRs can sometimes have cortical contributions (Coffey, Herholz, Chepesiuk, Baillet, & Zatorre, 2016) when the F0 of the stimulus is low (≤ 100 Hz). Above ~150 Hz, only subcortical (brainstem) sources contribute to the FFR (Bidelman, 2018). It is possible that at least some of the correlations between spectral properties of the FFR (e.g., F0) and various aspects of speech perception reported in earlier studies (Anderson et al., 2012; Samira Anderson et al., 2010b; Bidelman & Krishnan, 2010b; Coffey et al., 2017; Du, Kong, et al., 2011; A Parbery-Clark, Skoe, & Kraus, 2009) may be cortical, rather than *subcortical*, in origin. The lack of robust links between the FFR and concurrent speech perception in the present study may be due to the fact that our FFRs reflect more pre-attentive, exogenous neural encoding of the brainstem, which does not always covary with perceptual measures (Bidelman, Moreno, & Alain, 2013a; Gockel et al., 2011). While our data do not provide strong evidence that perceptual correlates of concurrent vowel processing exist in FFRs, brainstem signal processing is no doubt still critical in feeding

later decision-based mechanisms at a cortical level. That is, neural encoding in brainstem might ultimately enhance segregation and perception by higher-order cognitive processes (Bidelman & Alain, 2015a; Bidelman, Davis, & Pridgen, 2018). Concurrent recordings of FFR (brainstem) and ERP (cortical) responses could test this possibility.

Relationships between perceptual and brainstem auditory coding, where they do exist, can be viewed within the framework of corticofugal (top-down) tuning of sensory function. Corticofugal neural pathways, that project back to peripheral structures (Suga, Gao, Zhang, Ma, & Olsen, 2000; Zhang & Suga, 2005) may control and enhance subcortical encoding of the F0 (voice pitch)-and formant (vowel identity) related information of the stimulus that are necessary for speech-in-noise perception. Of the brain-behavior correlates we did observe, F1 was associated with behavioral RTs, particularly in noise. The higher variability in F1 responses may be due to greater individual differences in the encoding of these higher spectral cues in this more challenging listening condition and/or due to difficulty of the task—larger spreads which would allow for correlations. Alternatively, this variability may also be related to corticofugal tuning of sensory (FFR) encoding that enhances acoustic features of target speech subcortically (Anderson & Kraus, 2013; Reetzke, Xie, Llanos, & Chandrasekaran, 2018). In background noise, corticofugal functions might search for sensory features that allow the listener to extract and enhance pertinent speech information. This notion is consistent with previous neural data (Cunningham, Nicol, Zecker, Bradlow, & Kraus, 2001; Parbery-Clark, Marmel, Bair, & Kraus, 2011; Parbery-Clark, Skoe, Lam, & Kraus, 2009) and perceptual models showing changes in the weighting of perceptual dimensions because of feedback (Amitay, 2009; Nosofsky, 1987). Online corticofugal activity may adapt rapidly especially in challenging environments (e.g.,

noise) (Atiani, Elhilali, David, Fritz, & Shamma, 2009; Elhilali, Ma, Micheyl, Oxenham, & Shamma, 2009).

Still, why corticofugal effects would be present at F1 but not F0 is unclear. The corticofugal activity may be related to the change in the power of ongoing theta-band rhythms in noise, our previous work showed correspondence of theta-band activity with RTs in noise (Yellamsetty & Bidelman, 2018a). Thus, we anticipate the involvement of the lower oscillatory theta-rhythms in modulating the spectral feature at the subcortical level in noise. Moreover, our results are probably not due corticofugal mechanisms as we used a passive listening task whereas cortico-collicular efferents are thought to be recruited in tasks requiring goal-directed attention (Slee & David, 2015; Vollmer, Beitel, Schreiner, & Leake, 2017). Nevertheless, it would be interesting to see how the variable weighting of F0/F1 coding and simultaneous changes in oscillatory rhythms (specially theta-band) across individuals, in an active listening task. Attention might act to bias and enhance incoming acoustic speech relevant information and suppress noise (Suga, 2012).

A handful of studies have shown certain vowels dominate perception among different vowel pair combinations (Assmann & Summerfield, 1990a, 2004; Chintanpalli et al., 2014a, 2016; Chintanpalli & Heinz, 2013a; Meddis & Hewitt, 1992a), reminiscent of our vowel dominancy data (Fig. 5). At 0ST, listeners can take advantage of the relative differences in the levels of spectral peaks between two vowels and one vowel is identified dominantly over the other; whereas identification of both the vowels is better at 4 ST. Our stimulus did have a level difference between the spectral peaks (F1s) of the two vowels; /ɛ/ was slightly stronger (2 dB) than /a/ in acoustic power. This level difference is captured in FFR amplitudes (Fig. 7C). Indeed, when FFRs were split by listeners' behavior, double-vowel responses showed closer

correspondence to the single /ɛ/ vowel (Fig. 10). Thus, FFRs were largely independent of behavior bias and therefore showed a stimulus (rather than perceptual) dominancy.

In sum, we find that FFRs reflect the neuro-acoustic representations of peripheral nonlinearities that are carried forward to brainstem processing. The spectro-temporal changes observed in FFRs with pitch and noise cues and weak behavioral correlations suggest that FFRs reflect mainly exogenous stimulus properties of concurrent speech mixtures. Nevertheless, correlations between F1 and behavioral RTs in noisy listening conditions suggest possible corticofugal involvement in enhancing speech relevant representations in the brainstem during more difficult tasks and/or in challenging listening conditions. Our results show that FFRs reflect pre-attentive mechanisms and concurrent stimulus interactions that can, under certain conditions, predict the successful identification of speech mixtures.

Chapter 4

GENERAL DISCUSSION

Recording EEG at subcortical (pre-attentive) and cortical (post-attentive) levels for a concurrent speech stimulus illustrated the hierarchy of the processing underlying concurrent speech identification and elucidated the neural mechanics and time course of concurrent speech identification in clean and degraded listening conditions. Our results showed four primary findings (1) Behavioral results showed that listeners exploit F0-differences between vowels to identify speech, and the perceptual F0 benefits was larger for clean compared to noise degraded (+5 dB SNR) stimuli. (2) early in the auditory pathway, the pre-attentive mechanisms of concurrent stimulus interactions weakly predict the success of identifying simultaneous speech, indicating exogenous nature of the subcortical activity. (3) dynamic F0 cues and noise (SNR) are likely to interact during the extraction of multiple auditory streams and occur relatively early (~200 ms) in the neural hierarchy. (4) Higher band cortical rhythms carry information on pre-perceptual encoding of acoustic features and follow the quality of the speech signal, whereas low rhythms reflect post-perceptual cognitive operations that involve the degree of listening effort and task demands. In addition to this, we anticipate the involvement of lower oscillatory rhythms (especially theta-band) in modulating the spectral feature (F1) at the subcortical level in noise.

Human speech perception is based on multiple, hierarchical processing pathways, and different kinds of representations in speech could be preferentially treated in different streams (such as acoustic–phonetic features and articulatory gestures)(Scott & Johnsrude, 2003). When the signal reaches the ear, the acoustic wave is first decomposed into perceptual groups (i.e., source/objects) according to Gestalt principles (Koffka, 1935). In the current studies, double vowels with same ST (0ST) could be grouped together into one entity or stream because the

harmonics fall into the same auditory filter band channel. Whereas for different F0s (4ST), the harmonics fall across-channel filter bands forming two entities or streams. When the two signals represent simultaneous processes in the same processing stream (within channel), we would expect enhancements (Coffey et al., 2016), that is reflected in FFR-F0 at subcortical level and to be paralleled by enhancements in the strength of the N1-P2 component and γ_{high} activity at the cortical level. These enhancements are mere exogenous acoustic feature representations and larger amplitudes do not necessary show perceptual benefit as shown in the behavioral data. Whereas, when two signals are processed in two different streams (4ST), the energy gets distributed across the channels and carried as two separate sources to the higher cortical levels, this leads to better perceptual identification of the two speech sounds.

It has been known that noise tends to obscure amplitude modulations in speech that are essential for its comprehension (Bidelman, 2016b; Shannon et al., 1995; Swaminathan & Heinz, 2012). Subcortically, FFR amplitudes were larger for single vowel and showed reduction in amplitudes of both the pitch and timbre cues with competing speech signal (double vowels) and in noise. At cortical level, in clean conditions, neural rhythms showed larger amplitudes of the γ_{band} (200ms) followed by the desynchrony in the β- (around ~250 to 350 ms) and θ- oscillations (~400 -500 ms) post-stimulus onset. Whereas in noise, there was reduced amplitude of γ_{band} followed by increased activity in the β- and θ- oscillations compared to clean conditions. Thus, the tuning of non-linear acoustic properties of the speech signals encoded at the subcortical level may be general to all sounds and speech-specific operations probably do not begin until the signal reaches the cerebral cortex (Scott & Johnsrude, 2003).

A meta-analysis of the perceptual correlates across our studies supports this notion. Figure 11 shows the magnitude of correlation observed in study one and two. The perceptual

correlations of accuracy were stronger with the cortical activity and were weaker with the subcortical activity (Fig 11A). For the reaction time (fig. 11B), cortical activity has greater contribution and minimal contributions from the subcortical activity. The correlation shown at the cortical level might be covertly driven by the band specific changes in the oscillatory spectral responses, reflecting the unique mechanisms of speech perception.

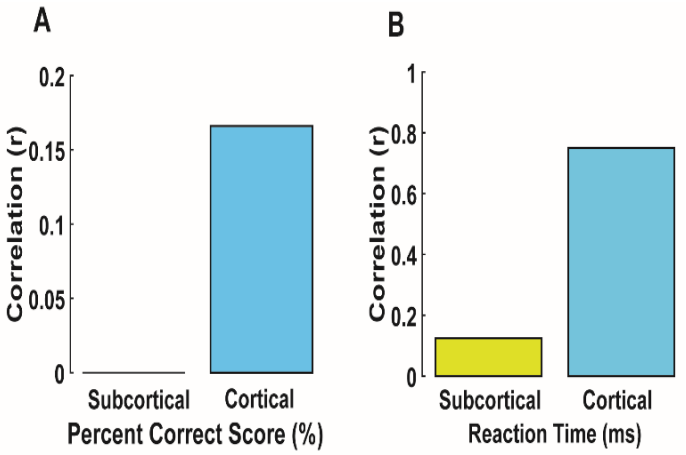

Figure 11. Meta-analysis across studies showing the distribution of the Brain-behavior correlates at cortical and subcortical levels for double-vowel perception. Bar plots showing the contribution of the subcortical and cortical levels (A) for the identification accuracy of concurrent double vowels and (B) for the behavioral reaction time.

Brainstem encoding of speech is a fundamental precursor to the divergence of the parallel processing streams identified in the cortex (Kraus & Nicol, 2005). Different qualities of the stimulus like periodicity (source) and spectra (filter) are processed separately, yet simultaneously (parallel sensory streams), by different neural mechanisms before the stimulus is consciously perceived as a whole (Kraus & Nicol, 2005). These separate source and filter characteristics are viewed independently in the response patterns of brainstem neurons and carried further in two streams of dorsal and ventral paths for localizing the source and identifying the object,

respectively (Kaas & Hackett, 1999; Rauschecker & Tian, 2000). The neurons in the primary and surrounding belt areas of auditory cortex are both sensitive to pitch and harmonically related frequencies (Bendor, Osmanski, & Wang, 2012b; Kikuchi, Horwitz, Mishkin, & Rauschecker, 2014a), whereas spectra in speech and mapping sound to meaning activates secondary auditory pathway regions (Kraus & Nicol, 2005; Saur et al., 2008). The two streams further integrate and contribute to functionally distinct regions of the frontal lobe (Romanski et al., 1999). In addition to this, A1 has both feedforward cortico-collicular and feedback cortico-cortical pathways, and this functional connectivity was a strong predictor of degraded speech perception (Bidelman et al., 2018). The cortical processes project backward to structures in the auditory periphery (Suga et al., 2000; Zhang & Suga, 2005) in case of speech in noise, these processes may enhance features of the target speech sounds subcortically (Anderson & Kraus, 2013; Reetzke et al., 2018). This explains the sequence of neural events and time course underlying the perception of concurrent speech identification in noise. The reduced γ-band activity and increase in the β-band and θ-band activity in noise would reflect the increase in the effort to extract phonetic features by enhancing the target speech features subcortically and perceptually binding them to match against the internalized memory profiles for both vowel constituents. Thus, we anticipate the involvement of the lower oscillatory rhythms (especially theta-band) in modulating the spectral features at the subcortical level in noise. Theoretically the cortical activity that modulates the brainstem encoding has reflected in the RT in noise, as seen in the fig. 11B. Hence, the pattern of distribution of reaction time correlations seen with F1 activity subcortically, and low frequency β- and θ- oscillatory activity cortically (Fig. 11B).

To summarize the brain-behavioral correlates of concurrent speech identification (fig 12), early pre-attentive subcortical activity transcribes the acoustic information and the actual

auditory percept occur at the cortical level. The cortical activity is an index of the reaction time that modulates the encoding of the spectral features at brainstem in noise.

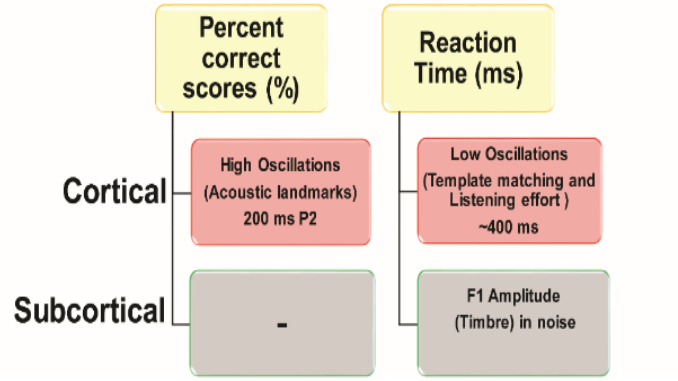

Figure 12. The hierarchical neural to behavioral correlates that are driving the identification of the concurrent double vowel stimulus.

To conclude, the dynamic time course for concurrent speech sound processing depends on both extrinsic (noise) and intrinsic (pitch) acoustic factors. Early in the auditory system, FFRs predicted perceptual speed but not the accuracy of double vowel identification, indicating the exogenous orientation of pre-attentive subcortical activity. Cortically, early high frequency activity reflected pre-perceptual encoding of acoustic features (~200 ms) and the quality (i.e., SNR) of the speech signal (~250-350ms), whereas later-evolving low-frequency rhythms (~400-500ms) reflected post-perceptual, cognitive operations that covaried with listening effort and task demands. Tentatively, we posit that successful speech identification is governed by peripheral Gestalt mechanics and cortical- accurate perceptual object construction, auditory template matching, and deceased listening effort/attentional allocation.

Future directions:

To test our anticipation of the lower oscillatory rhythms (specially theta-band) involvement in modulating the spectral feature (F1) at the subcortical level in noise. It would be interesting to localize the time course neural sources across the auditory pathway for double vowel identification process.

To test the above phenomenon on different target population like children with processing and learning difficulties, geriatric population, cochlear implant and hearing-impaired individuals. Extensions of this work, to examine how the weighting of neural activity changes across the pathway and on frequency bands with perceptual learning, attentional effects on concurrent speech segregation.

References

Aftanas, Varlamov, Pavlov, Makhnev, & Reva. (2001). Affective picture processing: event-related synchronization within individually defined human theta band is modulated by valence dimension. *Neuroscience letters, 303*(2), 115-118.

Aissani, Martinerie, Yahia-Cherif, Paradis, & Lorenceau. (2014). Beta, but not gamma, band oscillations index visual form-motion integration. *PloS one, 9*(4), e95541.

Alain. (2007a). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research, 229*(1-2), 225-236. doi: S0378-5955(07)00018-4 [pii]

10.1016/j.heares.2007.01.011

Alain, Arnott, & Picton. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance, 27*(5), 1072-1089.

Alain, Arsenault, Garami, Bidelman, & Snyder. (2017). Neural correlates of speech segregation based on formant frequencies of adjacent vowels. *Scientific Reports, 7*(40790), 1-11.

Alain, Reinke, He, Wang, & Lobaugh. (2005). Hearing two things at once: Neurophysiological indices of speech segregation and identification. *Journal of Cognitive Neuroscience, 17*(5), 811-818. doi: 10.1162/0898929053747621

Alain, Reinke, McDonald, Chau, Tam, Pacurar, & Graham. (2005). Left thalamo-cortical network implicated in successful speech separation and identification. *Neuroimage, 26*(2), 592-599. doi: 10.1016/j.neuroimage.2005.02.006

Alain, C., Snyder, J. S., He, Y., & Reinke, K. S. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cerebral Cortex, 17*(5), 1074-1084. doi: bhl018 [pii]

10.1093/cercor/bhl018

Alain, Claude. (2007b). Breaking the wave: effects of attention and learning on concurrent sound perception. *Hearing research, 229*(1), 225-236.

Alain, Claude, Snyder, Joel S, He, Yu, & Reinke, Karen S. (2006). Changes in auditory cortex parallel rapid perceptual learning. *Cerebral Cortex, 17*(5), 1074-1084.

Amitay. (2009). Forward and reverse hierarchies in auditory perceptual learning. *Learning & Perception, 1*(1), 59-68.

Anderson, & Kraus. (2010). Sensory-cognitive interaction in the neural encoding of speech in noise: a review. *Journal of the American Academy of Audiology, 21*(9), 575-585.

Anderson, & Kraus. (2013). *cABR: A neural probe of speech-in-noise processing.* Paper presented at the Proceedings of the International Symposium on Auditory and Audiological Research.

Anderson, Parbery-Clark, White-Schwoch, & Kraus. (2012). Aging affects neural precision of speech encoding. *Journal of Neuroscience, 32*(41), 14156-14164.

Anderson, Skoe, Chandrasekaran, Zecker, & Kraus. (2010a). Brainstem correlates of speech-in-noise perception in children. *Hearing research, 270*(1-2), 151-157.

Anderson, Samira, Skoe, Erika, Chandrasekaran, Bharath, Zecker, Steven, & Kraus, Nina. (2010b). Brainstem correlates of speech-in-noise perception in children. *Hearing research, 270*(1-2), 151-157.

Arehart, King, & McLean-Mudgett. (1997). Role of fundamental frequency differences in the perceptual separation of competing vowel sounds by listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research, 40*(6), 1434-1444.

Arnott, Grady, Hevenor, Graham, & Alain. (2005). The functional organization of auditory working memory as revealed by fMRI. *Journal of Cognitive Neuroscience, 17*(5), 819-831.

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *Neuroimage, 22*(1), 401-408. doi: 10.1016/j.neuroimage.2004.01.014

Assmann, & Summerfield. (1989a). Modeling the perception of concurrent vowels: vowels with the same fundamental frequency. *J Acoust Soc Am, 85*(1), 327-338.
Assmann, & Summerfield. (1990a). Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *Journal of the Acoustical Society of America, 88*(2), 680-697.
Assmann, & Summerfield. (1994a). The contribution of waveform interactions to the perception of concurrent vowels. *J Acoust Soc Am, 95*(1), 471-484.
Assmann, & Summerfield. (2004). The perception of speech under adverse conditions *Speech processing in the auditory system* (pp. 231-308): Springer.
Assmann, P. F., & Summerfield, Q. (1989b). Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency. *Journal of the Acoustical Society of America, 85*(1), 327-338.
Assmann, P. F., & Summerfield, Q. (1990b). Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *Journal of the Acoustical Society of America, 88*(2), 680-697.
Assmann, P. F., & Summerfield, Q. (1994b). The contribution of waveform interactions to the perception of concurrent vowels. *J Acoust Soc Am, 95*(1), 471-484.
Atiani, Elhilali, David, Fritz, & Shamma. (2009). Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron, 61*(3), 467-480.
Bashivan, Bidelman, & Yeasin. (2014). Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity. *European Journal of Neuroscience, 40*(12), 3774–3784. doi: 10.1111/ejn.12749
Bastiaansen, Van Der Linden, Ter Keurs, Dijkstra, & Hagoort. (2005). Theta responses are involved in lexical—Semantic retrieval during language processing. *Journal of cognitive neuroscience, 17*(3), 530-541.
Bendor, Osmanski, & Wang. (2012a). Dual-pitch processing mechanisms in primate auditory cortex. *Journal of Neuroscience, 32*(46), 16149 –16161.
Bendor, Osmanski, & Wang. (2012b). Dual-pitch processing mechanisms in primate auditory cortex. *Journal of Neuroscience, 32*(46), 16149-16161.
Bidelman. (2015a). Induced neural beta oscillations predict categorical speech perception abilities. *Brain Lang, 141*, 62-69. doi: 10.1016/j.bandl.2014.11.003
Bidelman. (2015b). Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. *Hearing research, 323*, 68-80.
Bidelman. (2016a). Relative contribution of envelope and fine structure to the subcortical encoding of noise-degraded speech. *The Journal of the Acoustical Society of America, 140*(4), EL358-EL363.
Bidelman. (2016b). Relative contribution of envelope and fine structure to the subcortical encoding of noise-degraded speech. *Journal of the Acoustical Society of America, 140*(4), EL358-363.
Bidelman. (2017a). Amplified induced neural oscillatory activity predicts musicians' benefits in categorical speech perception. *Neuroscience, 348*, 107–113. doi: 10.1016/j.neuroscience.2017.02.015
Bidelman. (2017b). Communicating in challenging environments: Noise and reverberation *The Frequency-Following Response* (pp. 193-224): Springer.
Bidelman, & Alain. (2015a). Hierarchical neurocomputations underlying concurrent sound segregation: Connecting periphery to percept. *Neuropsychologia, 68*, 38-50. doi: S0028-3932(14)00480-1 [pii]

10.1016/j.neuropsychologia.2014.12.020
Bidelman, & Alain. (2015b). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *Journal of Neuroscience, 35*(2), 1240 –1249.
Bidelman, Davis, & Pridgen. (2018). Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hearing Research.*

Bidelman, Gandour, & Krishnan. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and cognition, 77*(1), 1-10.
Bidelman, & Howell. (2016). Functional changes in inter- and intra-hemispheric auditory cortical processing underlying degraded speech perception. *Neuroimage, 124*, 581-590.
Bidelman, & Krishnan. (2010a). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain research, 1355*, 112-125.
Bidelman, & Krishnan. (2010b). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research, 1355*, 112-125.
Bidelman, Moreno, ., & Alain. (2013a). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage, 79*, 201-212.
Bidelman, Moreno, ., & Alain. (2013b). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage, 79*(1), 201-212. doi: S1053-8119(13)00452-7 [pii]

10.1016/j.neuroimage.2013.04.093
Bidelman, & Powers. (in press-a). Response properties of the human frequency-following response (FFR) to speech and nonspeech sounds: Level dependence, adaptation, and phase-locking limits. . *International Journal of Audiology.*
Bidelman, Villafuerte, Moreno, ., & Alain. (2014). Age-related changes in the subcortical–cortical encoding and categorical perception of speech. *Neurobiology of aging, 35*(11), 2526-2540.
Bidelman, & Walker. (2017). Attentional modulation and domain specificity underlying the neural organization of auditory categorical perception. *European Journal of Neuroscience, 45*, 690-699.
Bidelman, Weiss, M. W., Moreno, S., & Alain. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience, 40*(4), 2662-2673.
Bidelman, & Yellamsetty. (2017a). Noise and pitch interact during the cortical segregation of concurrent speech. *Hearing Research, 351*, 34-44. doi: 10.1016/j.heares.2017.05.008
Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *NeuroImage, 175*, 56–69.
Bidelman, G. M., & Powers, L. (in press-b). Response properties of the human frequency-following response (FFR) to speech and nonspeech sounds: Level dependence, adaptation, and phase-locking limits. *International Journal of Audiology*.
Bidelman, G. M., & Yellamsetty, A. . (2017b). Noise and pitch interact during the cortical segregation of concurrent speech. *Hearing Research, 351*, 34-44.
Bidet-Caulet, & Bertrand. (2009). Neurophysiological mechanisms involved in auditory perceptual organization. *Frontiers in Neuroscience, 3*(2), 182-191. doi: 10.3389/neuro.01.025.2009
Binder, Liebenthal, Possing, Medler, & Ward. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience, 7*(3), 295-301. doi: 10.1038/nn1198

nn1198 [pii]
Bregman. (1990a). *Auditory Scene Analysis*. Cambridge, MA: MIT.
Bregman. (1990b). Auditory Scene Analysis: The perceptual organization of sound. 1990: MIT Press, Cambridge, MA.
Brovelli, Ding, Ledberg, Chen, Nakamura, & Bressler. (2004). Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proceedings of the National Academy of Sciences of the United States of America, 101*(26), 9849-9854.
Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., Kawasaki, H., Steinschneider, M., & Howard, M. A., 3rd. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *Journal of Neurophysiology, 102*(4), 2358-2374. doi: 10.1152/jn.91346.2008
Buzsáki. (2005). Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus, 15*(7), 827-840.

Cariani, Peter A, & Delgutte, Bertrand. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology, 76*(3), 1698-1716.
Carlyon, Robert P. (2004). How the brain separates sounds. *Trends in cognitive sciences, 8*(10), 465-471.
Carney, Li, & McDonough. (2015). Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. *Eneuro, 2*(4), ENEURO. 0004-0015.2015.
Cedolin, Leonardo, & Delgutte, Bertrand. (2005). Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. *Journal of neurophysiology, 94*(1), 347-362.
Chandrasekaran, Hornickel, Skoe, Nicol, & Kraus. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. *Neuron, 64*(3), 311-319.
Chandrasekaran, B, & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology, 47*(2), 236-246.
Chintanpalli, Ahlstrom, & Dubno. (2014a). Computational model predictions of cues for concurrent vowel identification. *Journal of the Association for Research in Otolaryngology : JARO, 15*(5), 823-837. doi: 10.1007/s10162-014-0475-7
Chintanpalli, Ahlstrom, & Dubno. (2016). Effects of age and hearing loss on concurrent vowel identification. *Journal of the Acoustical Society of America, 140*(6), 4142. doi: 10.1121/1.4968781
Chintanpalli, & Heinz. (2013a). The use of confusion patterns to evaluate the neural basis for concurrent vowel identification. *Journal of the Acoustical Society of America, 134*(4), 2988-3000. doi: 10.1121/1.4820888
Chintanpalli, A., Ahlstrom, J. B., & Dubno, J. R. (2014b). Computational model predictions of cues for concurrent vowel identification. *Journal of the Association for Research in Otolaryngology, 15*(5), 823-837. doi: 10.1007/s10162-014-0475-7
Chintanpalli, Ananthakrishna, & Heinz, Michael G. (2013b). The use of confusion patterns to evaluate the neural basis for concurrent vowel identification a. *The Journal of the Acoustical Society of America, 134*(4), 2988-3000.
Chung, & Bidelman. (2016). Cortical encoding and neurophysiological tracking of intensity and pitch cues signaling English stress patterns in native and nonnative speakers. *Brain and Language, 155–156*, 49-57. doi: http://dx.doi.org/10.1016/j.bandl.2016.04.004
Coffey, Chepesiuk, Herholz, Baillet, & Zatorre. (2017). Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Frontiers in neuroscience, 11*, 479.
Coffey, Herholz, Chepesiuk, Baillet, & Zatorre. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature communications, 7*, 11070.
Culling. (1990). Exploring the conditions for the perceptual separation of concurrent voices using F0 differences. *Proc. Inst. Acoust, 12*, 559-566.
Culling, & Darwin. (1993). Perceptual separation of simultaneous vowels: Within and across-formant grouping by F 0. *The Journal of the Acoustical Society of America, 93*(6), 3454-3467.
Cunningham, Nicol, Zecker, Bradlow, & Kraus. (2001). Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clinical Neurophysiology, 112*(5), 758-767.
de Cheveigné, Kawahara, Tsuzaki, & Aikawa. (1997). Concurrent vowel identification. I. Effects of relative amplitude and F0 difference. *Journal of the Acoustical Society of America, 101*(5), 2839-2847. doi: doi:http://dx.doi.org/10.1121/1.418517
Delgutte, & Kiang. (1984). Speech coding in the auditory nerve: I. Vowel-like sounds. *The Journal of the Acoustical Society of America, 75*(3), 866-878.
Delorme, & Makeig. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21. doi: 10.1016/j.jneumeth.2003.10.009
Destexhe, Hughes, Rudolph, & Crunelli. (2007). Are corticothalamic 'up'states fragments of wakefulness? *Trends in neurosciences, 30*(7), 334-342.

Ding, Nai, & Simon, Jonathan Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences, 109*(29), 11854-11859.
Doelling, Arnal, Ghitza, & Poeppel. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage, 85*, 761-768.
Donoghue, Sanes, Hatsopoulos, & Gaál. (1998). Neural discharge and local field potential oscillations in primate motor cortex during voluntary movements. *Journal of neurophysiology, 79*(1), 159-173.
Du, He, Ross, Bardouille, Wu, Li, & Alain. (2010). Human auditory cortex activity shows additive effects of spectral and spatial cues during speech segregation. *Cerebral Cortex, 21*(3), 698-707.
Du, Kong, Wang, Wu, & Li. (2011). Auditory frequency-following response: A neurophysiological measure for studying the "cocktail-party problem". *Neuroscience & Biobehavioral Reviews, 35*(10), 2046-2057.
Du, Y., He, Y., Ross, B., Bardouille, T., Wu, X., Li, L., & Alain, C. (2011). Human auditory cortex activity shows additive effects of spectral and spatial cues during speech segregation. *Cereb Cortex, 21*(3), 698-707. doi: 10.1093/cercor/bhq136
Dyson, & Alain. (2004). Representation of concurrent acoustic objects in primary auditory cortex. *Journal of the Acoustical Society of America, 115*(1), 280-288.
Eisner, McGettigan, Faulkner, Rosen, & Scott. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience, 30*(21), 7179-7186. doi: 10.1523/jneurosci.4040-09.2010
Elhilali, Ma, Micheyl, Oxenham, & Shamma. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron, 61*(2), 317-329.
Fries, Reynolds, Rorie, & Desimone. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science, 291*(5508), 1560-1563.
Fujioka, Trainor, Large, & Ross. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *Journal of Neuroscience, 32*(5), 1791-1802.
Gardi, Merzenich, & McKean. (1979). Origins of the scalp-recorded frequency-following response in the cat. *Audiology, 18*(5), 353-380.
Ghitza. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in psychology, 2*, 130.
Ghitza. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Frontiers in psychology, 3*, 238.
Ghitza. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Frontiers in psychology, 4*, 138.
Ghitza, Giraud, & Poeppel. (2013). Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Frontiers in human neuroscience, 6*, 340.
Giraud, Kell, Thierfelder, Sterzer, Russ, Preibisch, & Kleinschmidt. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral cortex, 14*(3), 247-255.
Giraud, & Poeppel. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience, 15*(4), 511-517.
Glaser, Suter, Dasheiff, & Goldberg. (1976). The human frequency-following response: its behavior during continuous tone and tone burst stimulation. *Electroencephalography and Clinical Neurophysiology, 40*(1), 25-32.
Gockel, Carlyon, Mehta, & Plack. (2011). The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *Journal of the Association for Research in Otolaryngology, 12*(6), 767-782.
Goswami. (2011). A temporal sampling framework for developmental dyslexia. *Trends in cognitive sciences, 15*(1), 3-10.
Guthrie, & Buchwald. (1991). Significance testing of difference potentials. *Psychophysiology, 28*(2), 240-244.

Haarmann, Cameron, & Ruchkin. (2002). Neural synchronization mediates on-line sentence processing: EEG coherence evidence from filler-gap constructions. *Psychophysiology, 39*(6), 820-825.
Haenschel, Baldeweg, Croft, Whittington, & Gruzelier. (2000). Gamma and beta frequency oscillations in response to novel auditory stimuli: a comparison of human electroencephalogram (EEG) data with in vitro models. *Proceedings of the National Academy of Sciences, 97*(13), 7645-7650.
Hanslmayr, Gross, Klimesch, & Shapiro. (2011). The role of alpha oscillations in temporal attention. *Brain research reviews, 67*(1), 331-343.
Henry, Abrams, Forst, Mender, Neilans, Idrobo, & Carney. (2017). Midbrain synchrony to envelope structure supports behavioral sensitivity to single-formant vowel-like sounds in noise. *Journal of the Association for Research in Otolaryngology, 18*(1), 165-181.
Hornickel, Skoe, Nicol, Zecker, & Kraus. (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings of the National Academy of Sciences, 106*(31), 13022-13027.
Houtgast. (1974). Lateral suppression in hearing. *Acad. Pers BV, Amsterdam*.
Hutka, Alain, Binns, & Bidelman. (2013). Age-related differences in the sequential organization of speech sounds. *Journal of the Acoustical Society of America, 133* (6), 4177-4187.
Jane, J Yu, & Young, Eric D. (2000). Linear and nonlinear pathways of spectral information transmission in the cochlear nucleus. *Proceedings of the National Academy of Sciences, 97*(22), 11780-11786.
Kaas, & Hackett. (1999). 'What'and'where'processing in auditory cortex. *Nature neuroscience, 2*(12), 1045.
Keilson, Suzanne E, Richards, Virginia M, Wyman, Bradley T, & Young, Eric D. (1997). The representation of concurrent vowels in the cat anesthetized ventral cochlear nucleus: evidence for a periodicity-tagged spectral representation. *The Journal of the Acoustical Society of America, 102*(2), 1056-1071.
Kikuchi, Horwitz, Mishkin, & Rauschecker. (2014a). Processing of harmonics in the lateral belt of macaque auditory cortex. *Frontiers in neuroscience, 8*, 204.
Kikuchi, Horwitz, Mishkin, & Rauschecker. (2014b). Processing of harmonics in the lateral belt of macaque auditory cortex. *Frontiers in Neuroscience, 8*(10.3389/fnins.2014.00204), 1-13.
Killion, Niquette, Gudmundsen, Revit, & Banerjee. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 116*(4 Pt 1), 2395-2405.
Klatt. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*(3), 971-995.
Koffka. (1935). Principles of Gestalt Psychology, International Library of Psychology, Philosophy and Scientific Method: Harcourt Brace, New York.
Kozou, Kujala, Shtyrov, Toppila, Starck, Alku, & Naatanen. (2005). The effect of different noise types on the speech and non-speech elicited mismatch negativity. *Hearing Research, 199*(1-2), 31-39. doi: S0378-5955(04)00234-5 [pii]

10.1016/j.heares.2004.07.010
Kraus, & Nicol. (2005). Brainstem origins for cortical 'what'and 'where'pathways in the auditory system. *Trends in neurosciences, 28*(4), 176-181.
Krishnan. (1999). Human frequency-following responses to two-tone approximations of steady-state vowels. *Audiology and Neurotology, 4*(2), 95-103.
Krishnan. (2002a). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing research, 166*(1-2), 192-201.
Krishnan. (2002b). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing research, 166*(1), 192-201.
Krishnan, & Agrawal. (2010). Human frequency-following response to speech-like sounds: correlates of off-frequency masking. *Audiology and Neurotology, 15*(4), 221-228.

Krishnan, Bidelman, & Gandour. (2010). Neural representation of pitch salience in the human brainstem revealed by psychophysical and electrophysiological indices. *Hearing research, 268*(1), 60-66.
Krumbholz, Katrin, Eickhoff, Simon B, & Fink, Gereon R. (2007). Feature-and object-based attentional modulation in the human auditory "where" pathway. *Journal of Cognitive Neuroscience, 19*(10), 1721-1733.
Kujala, & Brattico. (2009). Detrimental noise effects on brain's speech functions. *Biological psychology, 81*(3), 135-143.
Lakatos, Karmos, Mehta, Ulbert, & Schroeder. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *science, 320*(5872), 110-113.
Li, & Jeng. (2011). Noise tolerance in human frequency-following responses to voice pitch. *The Journal of the Acoustical Society of America, 129*(1), EL21-EL26.
Liu, & Kewley-Port. (2004). Formant discrimination in noise for isolated vowels. *The Journal of the Acoustical Society of America, 116*(5), 3119-3129.
Liu, Liang-Fa, Palmer, Alan R, & Wallace, Mark N. (2006). Phase-locked responses to pure tones in the inferior colliculus. *Journal of neurophysiology, 95*(3), 1926-1935.
Luck. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA, USA: MIT Press.
Marsh, Brown, & Smith. (1974). Differential brainstem pathways for the conduction of auditory frequency-following responses. *Electroencephalography and clinical neurophysiology, 36*, 415-424.
McKeown. (1992). Perception of concurrent vowels: The effect of varying their relative level. *Speech Communication, 11*(1), 1-13.
Meddis, & Hewitt. (1992a). Modeling the identification of concurrent vowels with different fundamental frequencies. *J Acoust Soc Am, 91*(1), 233-245. doi: doi:http://dx.doi.org/10.1121/1.402767
Meddis, R., & Hewitt, M. J. (1992b). Modeling the identification of concurrent vowels with different fundamental frequencies. *Journal of the Acoustical Society of America, 91*(1), 233-245. doi: doi:http://dx.doi.org/10.1121/1.402767
Meddis, Ray, & Hewitt, Michael J. (1992c). Modeling the identification of concurrent vowels with different fundamental frequencies. *The Journal of the Acoustical Society of America, 91*(1), 233-245.
Miller, Schilling, Franck, & Young. (1997). Effects of acoustic trauma on the representation of the vowel/ε/in cat auditory nerve fibers. *The Journal of the Acoustical Society of America, 101*(6), 3602-3616.
Nilsson, Soli, & Sullivan. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95*(2), 1085-1099.
Nosofsky. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(1), 87.
Oldfield. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-113.
Oostenveld, & Praamstra. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology, 112*, 713-719.
Oxenham. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification, 12*(4), 316-331. doi: 10.1177/1084713808325881
Palmer. (1990a). The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers. *Journal of the Acoustical Society of America, 88*(3), 1412-1426.
Palmer. (1990b). The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear- nerve fibers. *The Journal of the Acoustical Society of America, 88*(3), 1412-1426.

Palmer, Alan R, & Winter, Ian M. (1992). Cochlear nerve and cochlear nucleus responses to the fundamental frequency of voiced speech sounds and harmonic complex tones. *Auditory Physiology and Perception, 83*, 231-239.
Palmer, AR. (1990c). The representation of the spectra and fundamental frequencies of steady-state single-and double- vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers. *The Journal of the Acoustical Society of America, 88*(3), 1412-1426.
Palva, Monto, Kulashekhar, & Palva. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proceedings of the National Academy of Sciences, 107*(16), 7580-7585.
Parbery-Clark, Marmel, Bair, & Kraus. (2011). What subcortical-cortical relationships tell us about processing speech in noise. *European Journal of Neuroscience, 33*(3), 549-557. doi: 10.1111/j.1460-9568.2010.07546.x
Parbery-Clark, Skoe, Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and hearing, 30*(6), 653-661.
Parbery-Clark, A, Skoe, E, & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience, 29*(45), 14100-14107.
Paus, Zatorre, Hofle, Caramanos, Gotman, Petrides, & Evans. (1997). Time-related changes in neural systems underlying attention and arousal during the performance of an auditory vigilance task. *Journal of cognitive neuroscience, 9*(3), 392-408.
Peters, Moore, & Baer. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America, 103*(1), 577-587.
Picton, Alain, Woods, John, Scherg, Valdes-Sosa, . . . Trujillo. (1999a). Intracerebral sources of human auditory-evoked potentials. *Audiology & Neuro-otology, 4*(2), 64-79. doi: 13823 [pii]
Picton, TW, Alain, C, Woods, DL, John, MS, Scherg, M, Valdes-Sosa, P, . . . Trujillo, NJ. (1999b). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neurotology, 4*(2), 64-79.
Pope, Kenneth J, Fitzgibbon, Sean P, Lewis, Trent W, Whitham, Emma M, & Willoughby, John O. (2009). Relation of gamma oscillations in scalp recordings to muscular activity. *Brain topography, 22*(1), 13-17.
Portfors, & Sinex. (2005). Coding of communication sounds in the inferior colliculus *the inferior colliculus* (pp. 411-425): Springer.
Prévost, Laroche, Marcoux, & Dajani. (2013). Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. *Clinical Neurophysiology, 124*(1), 52-60.
Pulvermüller. (1999). Words in the brain's language. *Behavioral and brain sciences, 22*(02), 253-279.
Rauschecker, & Tian. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences, 97*(22), 11800-11806.
Reale, & Geisler. (1980). Auditory-nerve fiber encoding of two- tone approximations to steady- state vowels. *The Journal of the Acoustical Society of America, 67*(3), 891-902.
Reetzke, Xie, Llanos, & Chandrasekaran. (2018). Tracing the Trajectory of Sensory Plasticity across Different Stages of Speech Learning in Adulthood. *Current Biology*.
Reinke, He, Wang, & Alain. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research, 17*, 781-791.
Romanski, Tian, Fritz, Mishkin, Goldman-Rakic, & Rauschecker. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature neuroscience, 2*(12), 1131.
Ross, Schneider, Snyder, & Alain. (2010). Biological markers of auditory gap detection in young, middle-aged, and older adults. *PloS One, 5*(4), e10101. doi: 10.1371/journal.pone.0010101

Ruggero, Robles, & Rich. (1992). Two-tone suppression in the basilar membrane of the cochlea: mechanical basis of auditory-nerve rate suppression. *Journal of neurophysiology, 68*(4), 1087-1099.
Russo, Nicol, Musacchia, & Kraus. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology, 115*(9), 2021-2030.
Sachs, & Kiang. (1968). Two-tone inhibition in auditory- nerve fibers. *The Journal of the Acoustical Society of America, 43*(5), 1120-1128.
Saur, Kreher, Schnell, Kümmerer, Kellmeyer, Vry, . . . Abel. (2008). Ventral and dorsal pathways for language. *Proceedings of the national academy of Sciences, 105*(46), 18035-18040.
Sauseng, Klimesch, Gruber, & Birbaumer. (2008). Cross-frequency phase synchronization: a brain mechanism of memory matching and attention. *Neuroimage, 40*(1), 308-317.
Scheffers. (1983). *Sifting vowels: Auditory pitch analysis and sound segregation.* Rijksuniversiteit te Groningen.
Scott, & Johnsrude. (2003). The neuroanatomical and functional organization of speech perception. *Trends in neurosciences, 26*(2), 100-107.
Shahin, Picton, & Miller. (2009). Brain oscillations during semantic evaluation of speech. *Brain and Cognition, 70*(3), 259-266. doi: 10.1016/j.bandc.2009.02.008
Shamma, Elhilali, & Micheyl. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences, 34*(3), 114-123. doi: 10.1016/j.tins.2010.11.002
Shannon. (1976). Two-tone unmasking and suppression in a forward- masking situation. *The Journal of the Acoustical Society of America, 59*(6), 1460-1470.
Shannon, Zeng, Kamath, Wygonski, & Ekelid. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303-304.
Shetty. (2016). Temporal cues and the effect of their enhancement on speech perception in older adults–A scoping review. *Journal of Otology, 11*(3), 95-101.
Shinn-Cunningham, & Best. (2008). Selective attention in normal and impaired hearing. *Trends in amplification, 12*(4), 283-299.
Sinex, & Geisler. (1983). Responses of auditory-nerve fibers to consonant –vowel syllables. *The Journal of the Acoustical Society of America, 73*(2), 602-615.
Sinex, Guzik, Li, & Sabes. (2003). Responses of auditory nerve fibers to harmonic and mistuned complex tones. *Hearing research, 182*(1), 130-139.
Sinex, Sabes, ., & Li. (2002a). Responses of inferior colliculus neurons to harmonic and mistuned complex tones. *Hearing Research, 168*, 150-162.
Sinex, Donal G. (2008). Responses of cochlear nucleus neurons to harmonic and mistuned complex tones. *Hearing research, 238*(1), 39-48.
Sinex, Donal G, Henderson, Jennifer, Li, Hongzhe, & Chen, Guang-Di. (2002). Responses of chinchilla inferior colliculus neurons to amplitude-modulated tones with different envelopes. *JARO-Journal of the Association for Research in Otolaryngology, 3*(4), 390-402.
Sinex, Donal G, Li, Hongzhe, & Velenovsky, David S. (2005). Prevalence of stereotypical responses to mistuned complex tones in the inferior colliculus. *Journal of neurophysiology, 94*(5), 3523-3537.
Sinex, Donal G, Sabes, Jennifer Henderson, & Li, Hongzhe. (2002b). Responses of inferior colliculus neurons to harmonic and mistuned complex tones. *Hearing research, 168*(1), 150-162.
Slee, S. J., & David, S. V. (2015). Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. *Journal of Neuroscience, 35*(38), 13090-13102. doi: 10.1523/JNEUROSCI.1671-15.2015
Smith, Marsh, & Brown. (1975). Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalography and clinical neurophysiology, 39*(5), 465-472.
Snyder, & Sinex. (2002). Immediate changes in tuning of inferior colliculus neurons following acute lesions of cat spiral ganglion. *Journal of neurophysiology, 87*(1), 434-452.
Song, Skoe, Banai, & Kraus. (2011). Perception of speech in noise: neural correlates. *Journal of cognitive neuroscience, 23*(9), 2268-2279.

Stillman, Crow, & Moushegian. (1978). Components of the frequency-following potential in man. *Electroencephalography and Clinical Neurophysiology, 44*(4), 438-446.
Studebaker. (1985). A "rationalized" arcsine transform. *Journal of Speech, Language, and Hearing Research, 28*(3), 455-462. doi: 10.1044/jshr.2803.455
Suga. (2012). Tuning shifts of the auditory system by corticocortical and corticofugal projections and conditioning. *Neuroscience & Biobehavioral Reviews, 36*(2), 969-988.
Suga, Gao, Zhang, Ma, & Olsen. (2000). The corticofugal system for hearing: recent progress. *Proceedings of the National Academy of Sciences, 97*(22), 11807-11814.
Swaminathan, & Heinz. (2012). Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *Journal of Neuroscience, 32*(5), 1747-1756. doi: 32/5/1747 [pii]

10.1523/JNEUROSCI.4493-11.2012
Tadel, Baillet, Mosher, Pantazis, & Leahy. (2011a). Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational intelligence and neuroscience, 2011*, 8.
Tadel, Baillet, Mosher, Pantazis, & Leahy. (2011b). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience, 2011*, 1-13.
Tallon-Baudry, & Bertrand. (1999a). Oscillatory gamma activity in humans and its role in object representation. *Trends in cognitive sciences, 3*(4), 151-162.
Tallon-Baudry, C., & Bertrand, O. (1999b). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences, 3*(4), 151-162.
Tan, & Carney. (2005). Encoding of vowel-like sounds in the auditory nerve: Model predictions of discrimination performance. *The Journal of the Acoustical Society of America, 117*(3), 1210-1222.
Trainor, Shahin, & Roberts. (2009). Understanding the benefits of musical training: Effects on oscillatory brain activity. *Annals of the New York Academy of Sciences, 1169*, 133-142. doi: 10.1111/j.1749-6632.2009.04589.x
Van Noorden. (1975). *Temporal coherence in the perception of tone sequences.* Doctoral Dissertation. Eindhoven University of Technology, Eindhoven, The Netherlands.
Vollmer, M., Beitel, R. E., Schreiner, C. E., & Leake, P. A. (2017). Passive stimulation and behavioral training differentially transform temporal processing in the inferior colliculus and primary auditory cortex. *Journal of Neurophysiology, 117*(1), 47-64. doi: 10.1152/jn.00392.2016
von Stein, & Sarnthein. (2000). Different frequencies for different scales of cortical integration: From local gamma to long range alpha/theta synchronization. *International Journal of Psychophysiology, 38*, 301-313.
Wallstrom, Kass, Miller, Cohn, & Fox. (2004). Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods. *International Journal of Psychophysiology, 53*, 105-119.
Worden, & Marsh. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalography and clinical neurophysiology, 25*(1), 42-52.
Xie, Reetzke, & Chandrasekaran. (2017). Stability and plasticity in neural encoding of linguistically relevant pitch patterns. *Journal of neurophysiology, 117*(3), 1409-1424.
Yellamsetty, & Bidelman. (2018a). Low-and high-frequency cortical brain oscillations reflect dissociable mechanisms of concurrent speech segregation in noise. *Hearing research, 361*, 92-102.
Yellamsetty, A., & Bidelman, G. M. (2018b). Low- and high-frequency cortical brain oscillations reflect dissociable mechanisms of concurrent speech segregation in noise. *Hearing Research, 361*, 92-102. doi: https://doi.org/10.1016/j.heares.2018.01.006
Young, & Sachs. (1979a). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *The Journal of the Acoustical Society of America, 66*(5), 1381-1403.

Young, Eric D, & Sachs, Murray B. (1979b). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *The Journal of the Acoustical Society of America, 66*(5), 1381-1403.

Zhang, & Suga. (2005). Corticofugal feedback for collicular plasticity evoked by electric stimulation of the inferior colliculus. *Journal of Neurophysiology, 94*(4), 2676-2682.

Zhou, Melloni, Poeppel, & Ding. (2016). Interpretations of Frequency Domain Analyses of Neural Entrainment: Periodicity, Fundamental Frequency, and Harmonics. *Frontiers in Human Neuroscience, 10*.

Zwicker. (1984). Auditory recognition of diotic and dichotic vowel pairs. *Speech Communication, 3*(4), 265-277.

Appendix

IRB Approval

Institutional Review Board
Office of Sponsored Programs
University of Memphis
315 Admin Bldg
Memphis, TN 38152-3370

PI: Gavin Bidelman
Co-Investigator:
Advisor and/or Co-PI:
Department: Institute For Intelligent Sys
Study Title: Neural correlates of complex auditory perception
IRB ID: 2370
Submission Type: Renewal
Level of Review: Expedited

IRB Meeting Date:
Decision: Approved
Approval Date: Jan 6, 2017
Expiration Date: Jan 6, 2018