ARTICLE IN PRESS

Hearing Research xxx (2018) 1–11

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

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

Anusha Yellamsetty [a], Gavin M. Bidelman [a, b, c, *]

[a] School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA
[b] Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
[c] Univeristy of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, TN, USA

ARTICLE INFO

Article history:
Received 24 May 2017
Received in revised form
9 December 2017
Accepted 12 January 2018
Available online xxx

Keywords:
EEG
Time-frequency analysis
Double-vowel segregation
F0-benefit
Speech-in-noise perception

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 (+5 dB 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 increments 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 are consistent with the notion 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.

1. 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 and Summerfield, 1989; Assmann and Summerfield, 1990; Assmann and Summerfield, 1994; de Cheveigné et al., 1997). This improvement has been referred to as the "F0 benefit" (Arehart et al., 1997; Chintanpalli and Heinz, 2013; Chintanpalli et al., 2016). 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 (Palmer, 1990; Portfors and Sinex, 2005; Sinex et al., 2003; Snyder and Sinex, 2002) and in human, via evoked potentials (Alain et al., 2005; Bidelman, 2017; Bidelman and Alain, 2015b; Dyson and Alain, 2004) and fMRI (Arnott et al., 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, 1990)] and phonetic template matching (Alain

---

[*] Corresponding author. School of Communication Sciences & Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA.
E-mail address: g.bidelman@memphis.edu (G.M. Bidelman).

https://doi.org/10.1016/j.heares.2018.01.006
0378-5955/© 2018 Elsevier B.V. All rights reserved.

Please cite this article in press as: Yellamsetty, A., Bidelman, G.M., Low- and high-frequency cortical brain oscillations reflect dissociable mechanisms of concurrent speech segregation in noise, Hearing Research (2018), https://doi.org/10.1016/j.heares.2018.01.006

et al., 2005; Meddis and Hewitt, 1992).

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., 2007, 2005, 2017; Bidelman and Yellamsetty, 2017; 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 (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 et al., 2005).

One of the main factors affecting the parsing of simultaneous speech is signal-to-noise ratio (SNR). 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 and Alain, 2015b; Bidelman and Howell, 2016; Binder et al., 2004; Eisner et al., 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, 2017; Bidelman and Yellamsetty, 2017).

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 et al., 2008), navigation (Buzsáki, 2005), memory (Palva et al., 2010; Sauseng et al., 2008), motor planning (Donoghue et al., 1998), and speech-language comprehension (Doelling et al., 2014; Ghitza, 2011, 2013; Ghitza et al., 2013; Haarmann et al., 2002; Shahin et al., 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 and 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, 2016). 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 (<3 Hz) 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 et al., 2005; Giraud and Poeppel, 2012; Goswami, 2011) and attention/arousal (Aftanas et al., 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, 2017; Fujioka et al., 2012; Ghitza, 2011), template matching (Bidelman, 2015a), lexical semantic memory access (Shahin et al., 2009), and perceptual binding in brain networks (Aissani et al., 2014; Brovelli et al., 2004; von Stein and Sarnthein, 2000). Lastly, gamma (γ: > 50 Hz) band has been associated with detailed phonetic features (Goswami, 2011), short duration cues (Giraud and Poeppel, 2012; Zhou et al., 2016), local network synchronization (Giraud and Poeppel, 2012; Haenschel et al., 2000), perceptual object construction (Tallon-Baudry and Bertrand, 1999), and experience-dependent enhancements in speech processing (Bidelman, 2017). 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). 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 (Alain et al., 2007; Bidelman and Yellamsetty, 2017; 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).

2. Methods

2.1. 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.

2.2. General speech-in-noise recognition task

We measured listeners' speech-in-noise (SIN) recognition using the standardized QuickSIN test (Killion et al., 2004). We have previously shown a strong correspondence between QuickSIN scores

Please cite this article in press as: Yellamsetty, A., Bidelman, G.M., Low- and high-frequency cortical brain oscillations reflect dissociable mechanisms of concurrent speech segregation in noise, Hearing Research (2018), https://doi.org/10.1016/j.heares.2018.01.006