Functional changes in inter- and intra-hemispheric cortical processing underlying degraded speech perception

Gavin M. Bidelman [a,b,*], Megan Howell [b]

[a] Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
[b] School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA

ABSTRACT

Previous studies suggest that at poorer signal-to-noise ratios (SNRs), auditory cortical event-related potentials are weakened, prolonged, and show a shift in the functional lateralization of cerebral processing from left to right hemisphere. Increased right hemisphere involvement during speech-in-noise (SIN) processing may reflect the recruitment of additional brain resources to aid speech recognition or alternatively, the progressive loss of involvement from left linguistic brain areas as speech becomes more impoverished (i.e., nonspeech-like). To better elucidate the brain basis of SIN perception, we recorded neuroelectric activity in normal hearing listeners to speech sounds presented at various SNRs. Behaviorally, listeners obtained superior SIN performance for speech presented to the right compared to the left ear (i.e., right ear advantage). Source analysis of neural data assessed the relative contribution of region-specific neural generators (linguistic and auditory brain areas) to SIN processing. We found that left inferior frontal brain areas (e.g., Broca's areas) partially disengage at poorer SNRs but responses do not right lateralize with increasing noise. In contrast, auditory sources showed more re-silience to noise in left compared to right primary auditory cortex but also a progressive shift in dominance from left to right hemisphere at lower SNRs. Region- and ear-specific correlations revealed that listeners' right ear SIN advantage was predicted by source activity emitted from inferior frontal gyrus (but not primary auditory cortex). Our findings demonstrate changes in the functional asymmetry of cortical speech processing during adverse acoustic conditions and suggest that "cocktail party" listening skills depend on the quality of speech representations in the left cerebral hemisphere rather than compensatory recruitment of right hemisphere mechanisms.

1. Introduction

Speech communication rarely occurs in quiet environments as nearly all real-world listening situations (e.g., classrooms, cocktail parties, restaurants) contain some degree of noise interference (Helfer and Wilber, 1990). Extracting relevant information from undesirable auditory scenes is hindered by additional competing sounds to target speech. Indeed, language and literacy skills can be compromised when learning in noisy environments (e.g., Bronzaft, 2002). Additive noise acts as a simultaneous masker, obscuring less intense portions of the speech signal and reducing its signal-to-noise ratio (SNR). Reduced SNR prevents audible access to salient speech cues (e.g., temporal envelope) normally exploited for robust comprehension (Shannon et al., 1995).

Noise exclusion deficits are magnified with age and hearing loss (Harris and Swenson, 1990; Hazrati and Loizou, 2012; Nabelek, 1988). Yet, current hearing aids provide little benefit for speech-in-noise (SIN) understanding despite restoring audiometric thresholds (Chmiel

and Jerger, 1996). It is now well accepted that SIN perception cannot be reliably predicted from the audiogram (Killion and Niquette, 2000). Moreover, SIN perception is problematic and highly variable among individuals without substantial hearing impairment (Divenyi and Haupt, 1997; Frisina and Frisina, 1997; Middelweerd et al., 1990) and even normal-hearing young adults (Song et al., 2011, 2012). These findings challenge conventional and longstanding views that speech intelligibility is determined solely by audibility, i.e., peripheral hearing status (Humes and Christopherson, 1991; Plomp, 1986; van Rooij et al., 1989). Rather, hearing sensitivity alone is inadequate to account for SIN perception issues (Humes and Christopherson, 1991; Parbery-Clark et al., 2011). Consequently, a growing body of evidence suggests that central auditory processing plays a critical role in mediating robust perceptual SIN abilities.

Auditory event-related brain potentials (ERPs) offer a precise temporal window to understand how noise affects the neural representation for speech and how central auditory brain mechanisms influence SIN listening skills. Noise-induced changes in the magnitude and timing of the auditory cortical ERPs have been reported by comparing responses to clean relative to noise-degraded speech sounds. The cortical encoding of auditory stimuli amidst noise reflects a complex interaction

* Corresponding author at: School of Communication Sciences & Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA. Fax: +1 901 525 1282.
E-mail address: g.bidelman@memphis.edu (G.M. Bidelman).

between the types of signal/noise, as well as the evoking stimulus paradigm (e.g., sequential vs. oddball paradigm) (Billings et al., 2010). Nevertheless, component waves of the ERPs can be suppressed (i.e., delayed and reduced in amplitude) (Baltzell and Billings, 2014; Billings et al., 2009, 2010) or facilitated (i.e., enhanced in amplitude) (Alain et al., 2009; Parbery-Clark et al., 2011) depending on the type (e.g., white noise, multi-talker babble) and effectiveness of a concurrent noise in masking the target signal. Importantly, behavioral SIN skills are directly related to the magnitude of these noise-related changes in neural activity (Bennett et al., 2012; Bidelman and Dexter, 2015; Billings et al., 2013; Parbery-Clark et al., 2011). Collectively, these studies demonstrate that early cortical neural representations are sensitive to the SNR of the speech signal. More critically, they suggest that noise inhibits the robust encoding of speech acoustics, resulting in the delivery of impoverished neural representation(s) to perceptual mechanisms operating downstream.

Auditory scalp-recorded potentials reflect the engagement of multiple brain networks overlapping in both space and time. As such, it is difficult to ascribe noise-related changes in a particular ERP deflection to a single neural generator. Nevertheless, gross changes in cerebral activation and functional asymmetry (i.e., hemispheric weighting) have been reported during SIN perception. Under normal circumstances, the auditory system shows a prominent leftward lateralization for speech processing, consistent with the well-known functional bias and left hemisphere (LH) specialization for linguistic functions (Tervaniemi and Hugdahl, 2003; Zatorre et al., 1992). In noisy listening conditions however, ERPs show a progressive increase in rightward activation. This reallocation in neural activity has been interpreted as suggesting that right hemisphere (RH) brain areas are recruited to aid degraded speech recognition (e.g., Shtyrov et al., 1998, 1999). However, it is unclear from previous studies if this enhancement in RH activity reflects additional compensatory processing to assist SIN understanding (e.g., Bidelman and Dexter, 2015; Du et al., 2014; Wong et al., 2009) or alternatively, a loss of linguistic function and hence residual engagement of resources that are more specialized to process non-speech sounds (e.g., Bidelman and Dexter, 2015; Zendel et al., 2015).

To further elucidate the neural basis of SIN listening, we recorded neuroelectric activity in young adult listeners while listening to speech sounds presented in ongoing noise at various SNRs. We applied a distributed source analysis to ERP responses to evaluate region-specific source generator differences and lateralization for the neural encoding of acoustically-impoverished speech. Comparing listeners' electrical brain responses to their behavioral performance allowed us to directly assess the degree to which isolated neural substrates (auditory vs. linguistic brain areas) contribute to behavioral SIN abilities. Consistent with previous reports, we hypothesized that noise would both weaken and prolong auditory cortical responses (e.g., Billings et al., 2009, 2010) and modulate functional lateralization (e.g., Shtyrov et al., 1998) dependent on the speech SNR. However, extending previous findings, we predicted that SIN perception would decline concomitant with diminished neural activity in either linguistic or auditory brain areas in the left hemisphere. This finding would support the notion that SIN perception is primarily driven not by an enhancement (i.e., compensation) of neural processing from RH, per se, but rather, a loss in quality of neural representation within (left) linguistic brain regions.

2. Methods

2.1. Participants

Twelve, young adults (mean ± SD age: 24.7 ± 2.7 years) participated in the experiment. All had obtained a similar level of formal education (at least a collegiate-level undergraduate degree) and were monolingual speakers of American English. Musical training is known to amplify the auditory evoked potentials (e.g., Bidelman et al., 2011; Musacchia et al., 2008; Zendel and Alain, 2009) and improve SIN listening skills

(Bidelman and Krishnan, 2010; Parbery-Clark et al., 2009; Zendel et al., 2015). Hence, all participants were required to have minimal formal musical training (1.3 ± 1.8 years) and none within the past five years. Air conduction audiograms confirmed normal hearing thresholds (i.e., ≤25 dB HL) at octave frequencies between 250 and 8000 Hz. Subjects also reported no history of hearing or neuropsychiatric disorders. Each gave informed written consent in compliance with a protocol approved by the University of Memphis Institutional Review Board and were reimbursed monetarily for their time.

2.2. Stimuli

Cortical auditory ERPs were elicited by a 300 ms/vCv/speech token /ama/ (cf. Bidelman, 2015; Shannon et al., 1999) (Fig. 1). The stimulus was a natural production recorded by a male speaker. The 50 ms nasal (/m/) was flanked by each vowel phoneme (/a/), both 125 ms in duration. The pitch prosody fell gradually over the duration of the token from an F0 of 120 Hz to 88 Hz. Vowel formant frequencies (F1–F3) were 830, 1200, and 2760 Hz, respectively. The intensity of the token was relatively fixed across its time course. In addition to this no noise "clean" stimulus (SNR = +∞ dB), noise-degraded speech stimuli were created by adding multitalker noise babble (Nilsson et al., 1994) to the clean token at SNRs of +10 and +5 dB. Importantly, SNR was manipulated by changing the level of the masker rather than the level of the signal. This ensured that SNR was inversely correlated with overall sound level (Binder et al., 2004). The babble was presented continuously throughout the (noise) experimental runs (i.e., the noise was not time-locked to the stimulus presentation) and was initiated ~5 s prior to delivery of the target speech stimuli. Continuous noise more closely mimics real-world situations whereby a listener is faced with extracting target signals above a blanket of competing background interference (e.g., cocktail party scenario) (e.g., Alain et al., 2012). The use of babble is also desirable as it has a similar deleterious effect on speech perception as other forms of interference (e.g., white noise) but a larger effect on neural encoding (Kozou et al., 2005).

2.3. Behavioral speech-in-noise task

We measured listeners' speech reception thresholds in noise using the QuickSIN test (Killion et al., 2004). The QuickSIN provides an efficient means to measure noise-degraded speech understanding and provides a standardized behavioral measure of SIN listening skills. In the present study, participants were presented with two lists of six sentences with five keywords per sentence embedded in four-talker
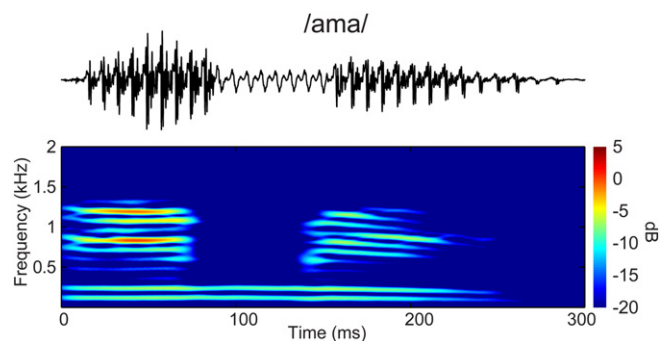

Fig. 1. Speech stimulus used to elicit cortical ERPs. Top, time waveforms of the vCv token /ama/; bottom, spectrogram. Tokens were 300 ms based on natural production of a male speaker. The 50 ms nasal (/m/) was flanked by each vowel phoneme (/a/), both 125 ms in duration. The pitch fell gradually over the duration of the token from an F0 of 120 Hz to 88 Hz. Vowel formant frequencies (F1–F3) were fixed at 830, 1200, and 2760 Hz, respectively. Noise babble was parametrically added to this clean token to achieve SNRs of +10 and +5 dB.