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

A R T I C L E   I N F O

Article history:
Received 6 March 2015
Accepted 9 September 2015
Available online 16 September 2015

Keywords:
Auditory scene analysis
Event-related brain potentials (ERPs)
Speech processing
Hemispheric laterality
Neural compensation
Reverse hierarchy theory (RHT)
Speech-in-noise (SIN) perception

A B S T R A C T

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 resilience 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.

© 2015 Elsevier Inc. All rights reserved.

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...
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 activation. This reallocation in neural activity has been interpreted as suggesting that right hemisphere (RH) brain areas are recruited to aid downstream.

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. Noise babble was parametrically added to this clean token to achieve SNRs of +10 and +5 dB.

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
were then digitally (2004). The PCA decomposition provided a set of independent components (saccades and blink artifacts) were then corrected in the continuous recording, excessively noisy channels were interpolated. Ocular artifacts from the continuous EEG traces to nullify ocular contamination in the line to a common average reference. Contact impedances were main-

be obtained below 5 kΩ. Horizontal EOGs were acquired from each eye using a bipolar method with Ag/AgCl electrodes placed ~1 cm posterior to Cz. However, data were re-referenced off-acquisition, all electrodes were referenced to an additional sensor coupler (IEC 60126). Left and right ear channels were calibrated (fixed, rarefaction polarity. SNR condition order was randomized within and across participants. The interstimulus interval (i.e., stimulus offset-to-onset) was jittered randomly between 480 and 720 ms (20-ms steps, rectangular distribution) to avoid rhythmic entrainment of the EEG (Luck, 2005, p. 168). The choice of these parameters ensured a balance between response recording time and minimizing stimulus-specific refractory (Picton et al., 1978) and component habitation effects (Picton et al., 1977) inherent to cortical auditory ERPs.

Neuroelectric activity was recorded from 64 sintered Ag/AgCl electrodes at standard 10-10 locations around the scalp (Oostenveld and Praamstra, 2001). EEGs were digitized using a sampling rate of 5000 Hz (SynAmps RT amplifiers; Compumedics Neuroscan). This high sampling frequency was necessary to record the extended bandwidth of the brainstem frequency-following response (not reported herein). Responses were then stored to disk for offline analysis. Electrodes placed on the outer canthi of the eyes and the superior and inferior orbit were used to monitor ocular activity. During online acquisition, all electrodes were referenced to an additional sensor placed ~1 cm posterior to Cz. However, data were re-referenced off-line to a common average reference. Contact impedances were maintained below 5 kΩ throughout the duration of the experiment. EEGs were then digitally filtered (1.5–20 Hz, zero-phase) for response visualization and quantification.

Subsequent preprocessing was performed in Curry 7 (Compumedics Neuroscan) and custom routines coded in MATLAB. Data visualization and scalp topographies were computed using EEG/ERPLAB (Delorme and Makeig, 2004; Lopez-Calderon and Luck, 2014). Prior to artifact correction, excessively noisy channels were interpolated. Ocular artifacts (saccades and blink artifacts) were then corrected in the continuous EEG using a principal component analysis (PCA) (Wallstrom et al., 2004). The PCA decomposition provided a set of independent components which best explained the topography of the blink/saccadic artifacts. The scalp projection of the first two PCA loadings was subtracted from the continuous EEG traces to nullify ocular contamination in the final ERPs. Cleaned EEGs were then epoched (−200–550 ms), baseline-corrected (i.e., the mean pre-stimulus voltage was subtracted from each epoch time point), and subsequently averaged in the time domain to obtain ERPs for each SNR condition per participant. In total, ERPs reflect the average of 2000 presentations of each stimulus condition. The entire experimental protocol including behavioral and electrophysiological testing took ~2 h to complete.

2.4.2. ERP source analysis

Neuronal sources of evoked potentials must be inferred given the volume-conducted nature of the scalp-recorded EEG and “cross-talk” between adjacent sensor measurements. To more directly assess generator characteristics underlying speech-evoked ERPs and SIN processing we performed a distributed source analysis (e.g., Bidelman and Dexter, 2015). In this study, a distributed source approach is preferable to a simpler paired dipole model as we expected foci to vary as a function of speech SNR (e.g., Shtryov et al., 1999); changes in location with stimulus SNR would render direct comparison of time-varying source strength spurious, as they would reflect signals generated in different brain areas. A distributed source approach allowed us to examine detailed neural activation within fixed regions of interest (ROIs). Source reconstruction was implemented in the MATLAB package Brainstorm (Tadel et al., 2011). We used a realistic, boundary element model (BEM) volume conductor (Fuchs et al., 1998, 2002) standardized to the MNI brain (Mazziotta et al., 1995). A BEM is less prone to spatial errors than other head models (e.g., concentric spherical conductor) (Fuchs et al., 2002).

The well-established sLORETA inverse solution (Pascual-Marqui, 2002) was used to estimate the distributed neuronal current density underlying the recorded sensor data. This algorithm models the inverse solution as a large number of elementary dipole generators distributed over nodes on a mesh of the cortical surface. When constrained to neocortical layers, the aggregate strength of source activity can be projected spatiotemporally onto the neuroanatomy, akin to functional maps in fMRI. The resultant activation maps represent the transcranial current source density underlying the scalp-recorded potentials as seen from the cortical surface. We used the default settings in Brainstorm's implementation of sLORETA (Tadel et al., 2011). From each sLORETA map, we extracted the time-course of source activity within two predefined ROIs: (1) bilateral primary auditory cortex (A1) (i.e., Heschl's gyrus, HG), (2) bilateral insula situated in inferior frontal gyrus (IFG). These ROIs were chosen to directly contrast source activity in brain regions subserving lower-order auditory processing and higher-order linguistic functions (e.g., Broca's area); they also allowed us to compare current results to those of other recent studies using identical ROIs (e.g., Bidelman and Dexter, 2015; Du et al., 2014). ROI parcellation was based on anatomical segmentations defined in FreeSurfer's Desikan–Killiany Atlas (Desikan et al., 2006) and the OpenMEEG BEM head model (Gramfort et al., 2010) as implemented in Brainstorm (Tadel et al., 2011)—see also, Fig. 1 of Klein and Tourville (2012). Resulting source waveforms reflect the neural activity (current, measured in μA/mm) as seen within each anatomical ROI.

Peak amplitude and latency were measured for the prominent deflections of the source waveforms (P1, N1, P2, N2) in specific time windows. Following conventions in previous studies (e.g., Bidelman et al., 2013; Irimajiri et al., 2005), P1 was defined as the maximum positive wave between 50 and 80 ms, N1 the negative-going trough between 85 and 150 ms, P2 as the positive-going peak between 150–250 ms, and N2 as the negativity between 220 and 350 ms. All latency measures were corrected for the acoustic delay of the headphone transducer.

2.4.3. Laterality of auditory cortical source responses

Changes in hemispheric laterality of speech processing with additive noise were assessed by comparing neuroelectric source activity from the left and the right hemispheres. A laterality index (%) was computed as LI := 100 × (R − L) / (R + L), where R and L are the N1 response magnitudes (i.e., [amplitude]) from left and right hemispheres, respectively.
(cf. King et al., 1999). Responses eliciting symmetric cortical activity produce a value of zero; asymmetric responses yield a positive or negative value for right or left hemisphere dominance, respectively.

2.5. Statistical analysis

Unless otherwise specified, two-way, generalized linear mixed-model ANOVAs were conducted on all dependent variables (GLIMMIX Procedure, SAS® 9.4, SAS Institute, Inc., Cary, NC, USA). Stimulus SNR (3 levels; clean, +10 dB, +5 dB) and hemisphere (2 levels; LH, RH) functioned as the within-subjects factor; subjects served as a random factor. Tukey–Kramer multiple comparisons controlled Type I error inflation. Significance in the laterality of source activity was first assessed using t-tests against zero (i.e., null hypothesis of bilateral symmetric response). We then compared shifts in laterality with increasing noise levels using a one-way, mixed-model ANOVA. An a priori significance level was set at $\alpha = 0.05$ for all statistical analyses. Normality and homogeneity of variance assumptions were confirmed prior to statistical inference.

Correlational analyses were used to determine the extent to which cortical responses could predict behavioral SIN performance, as measured via the QuickSIN. Spearman rank correlations were employed as this measure is preferable for small sample sizes and is less prone to artifacts due to outliers. Bootstrap resampling (Efron and Tibshirani, 1993) was used to validate significant brain-behavior correlations and estimate 95% confidence intervals (CIs). Participants were randomly resampled (with replacement) $N = 1000$ times from the original dataset. From each bootstrap surrogate, the correlation between neural and behavioral measures was recomputed. The probability density distribution of these statistics enabled us to confirm reliability of the observed correlations and estimate their variance.

3. Results

3.1. Behavioral speech-in-noise performance

Behavioral SIN scores, as measured via the QuickSIN are shown in Fig. 2. In general, participants achieved 0.5–1 dB better recognition for speech delivered to the right as compared to the left ear [paired samples t-test (two-tailed): $t_{11} = 2.66, p = 0.022$]. Consistent with the well-known asymmetry for speech perception in normal (i.e., quiet) listening conditions (Hiscock and Kinsbourne, 2011; Kimura, 1961), this finding indicates a similar right-ear advantage for degraded speech recognition.

Fig. 2. Behavioral QuickSIN scores for left vs. right ear. Listeners achieved better SIN perception for speech delivered to the right as compared to the left consistent with the well-known right-ear advantage in (quiet) speech perception. * $p < 0.05$; errorbars $= \pm 1$ s.e.m.

3.2. Cortical ERPs during SIN processing

Scalp topographies and time waveforms of the cortical ERPs (electrode data) as function of noise level are shown in Fig. 3. Visual inspection of scalp maps suggested weaker responses in the positive deflections (e.g., P1, P2) but stronger responses in the negative deflections (e.g., N1, N2) with increasing noise level (i.e., decreasing SNR) (see also Fig. S2). Hemispheric laterality also shifted rightward with increasing noise (Fig. S3). See Supplemental Material for analysis of the scalp data. The volume-conducted nature of sensor-space (i.e., electrode) recordings did not allow us to separate the underlying sources that contribute to these apparent functional changes. Consequently, subsequent analyses were conducted in source space to directly assess the neural generator characteristics underlying SIN processing.

3.3. Source activity during SIN perception

Whole brain source activations for clean and noise-degraded speech processing are shown in Fig. 4. Maps represent the statistical contrast (t-value, paired test) of sLORETA neural activations between the clean and +5 dB SNR conditions at 115 ms (−N1) and 230 ms (−N2) after stimulus onset. Maps were masked for multiple comparisons via false-discovery rate (FDR) on the family of contrasts under investigation ($\alpha = 0.05$) (Benjamini and Hochberg, 1995) and projected onto the semi-inflated cortical surface of the MNI brain (Collins et al., 1998). As indicated by the N1, clean speech recruited auditory cortex (HG) bilaterally but also engaged linguistic brain regions in IFG (Insula/Broca’s area). This finding confirms a significant contribution of frontal brain areas to the scalp-recorded auditory evoked responses to speech (see also Fig. S3). Contrastively, we found that noise-degraded speech showed stronger (N2) responses in HG of the right hemisphere, consistent with a rightward lateralization observed in the topographies of electrode-level data. This apparent rightward shift in hemispheric laterality of the speech-evoked ERPs could represent a stronger recruitment of RH with noise-degradation (e.g., Fig. S3; Sh tysrov et al., 1998; Shtyrov et al., 1999) or a disengagement of linguistic brain regions with the progressive loss of intelligibility (e.g., Fig. 4; Bidelman and Dexter, 2015; Zendel et al., 2015). Analysis of ROI-specific source waveforms extracted from within HG and IFG was used to tease apart these two interpretations.

3.4. Auditory cortical source responses during SIN processing

Fig. 5 illustrates source waveforms extracted from left and right primary auditory cortex (Heschl’s gyrus). An ANOVA revealed a simple main effect of SNR on P1 [F$_{2,55} = 6.02, p = 0.0043$] and N2 amplitudes [F$_{2,55} = 8.05, p = 0.0009$] but no hemispheric effects (ps $> 0.05$). Paralleling sensor-space results (i.e., Fig. S2), source P1 decreased and N2 increased in magnitude for noisier speech. P2 amplitudes were invariant to SNR [F$_{2,55} = 2.16, p = 0.12$] and were similar across hemispheres [F$_{1,55} = 0.06, p = 0.44$]. In contrast to other waves, N1 amplitudes showed a significant SNR × hemisphere interaction [F$_{2,55} = 4.32, p = 0.0181$] (Fig. 5B). Follow-up contrasts revealed prominent hemispheric differences in the +10 dB SNR condition. That is, at even moderate noise levels, source activity elicited by noise-degraded speech was generally larger (i.e., more positive) in the left compared to the right hemisphere.

With regard to latency, P1 occurred earlier in the right compared to the left hemisphere [F$_{1,55} = 7.11, p = 0.01$]. Also consistent with electrode-level responses, N1 and P2 source latencies were modulated by SNR, being prolonged for degraded compared to clean speech [N1: F$_{2,55} = 6.21, p = 0.0037$; P2: F$_{2,55} = 47.20, p < 0.0001$]. Neither a SNR nor a hemisphere effect was observed for N2 latency. Collectively, these findings imply that (i) the neural encoding of speech within auditory cortex is less efficient with increasing noise (i.e., delayed...
Fig. 3. Scalp topographies and butterfly plot of ERP time waveforms (electrode-level data) as a function of SNR. Cortical responses appeared as a series of obligatory positive and negative deflections (P1-N1-P2-N2 “waves”) within the first ~150 ms after the time-locked stimulus. Response waves were modulated in both amplitude and latency with changes in speech SNR. Scalp maps illustrate voltage distributions at the latency of the P2 and N2 waves, respectively. Later and weaker responses with increasing noise suggest a progressive loss of efficiency in the early cortical processing of speech (see also Figs. S1 and S2). Arrow, onset of the time-locking speech token.

Fig. 4. Cortical source activations for clean and noise-degraded speech processing. Maps show FDR-corrected \( p < 0.05 \) activations at 115 ms \( \sim \text{N1} \) and 230 ms \( \sim \text{N2} \) after stimulus onset between the clean and +5 dB SNR conditions projected onto the semi-inflated cortical surface of the MNI brain (Collins et al., 1998). Hot colors: brain areas where clean speech > noisy speech; cool colors: noise > clean. Clean speech recruits auditory cortex (HG) bilaterally but also engages linguistic brain regions in IFG regions (Ins./Broca’s area). Contrastively, noise-degraded speech shows stronger responses in HG of the RH. HG, Heschl’s gyrus; IFG, inferior frontal gyrus; Ins, Insula; LH/RH, left/right hemisphere.
responses) and (ii) there is less noise-related change (i.e., more resilience) in response magnitude in left compared to right auditory cortex.

Functional laterality of SIN processing in auditory cortex is shown in Fig. 5C. Note that |LI| ≥ 20% is typically considered “lateralized” (Seghier, 2008). Quantitative analyses confirmed a shift in hemispheric dominance with increasing noise levels. Whereas clean speech was strongly lateralized in left hemisphere \( t_{11} = -2.11, p = 0.028 \), the introduction of noise produced a shift toward the right hemisphere. Specifically, laterality diminished (became more bilateral) at +10 dB SNR \( t_{11} = 0.75, p = 0.76 \) and began showing a strong rightward dominance at +5 dB SNR \( t_{11} = 2.45, p = 0.01 \). These observations were confirmed by an ANOVA, which revealed a significant effect of SNR on source laterality \( F_{2, 55} = 8.77, p = 0.0045 \) (Fig. 6B). The sole main effect of hemisphere indicates that the N1 in frontal sources was larger (i.e., more negative) in LH across the board. Laterality of the IFG N1 revealed that clean speech was marginally lateralized in the left hemisphere \( t_{11} = -1.68, p = 0.06 \) and significantly left lateralized with moderate noise (+10 dB SNR: \( t_{11} = -2.27, p = 0.02 \)). Responses became bilateral for the +5 dB SNR condition \( t_{11} = -1.07, p = 0.85 \).

3.5. Inferior frontal gyrus source responses during SIN processing

Fig. 6 illustrates source waveforms extracted from left and right IFG. IFG responses were more variable than in the A1 ROI; only the N1 wave was clearly identifiable for subsequent analysis (Fig. 6A). An ANOVA conducted on IFG N1 source amplitudes revealed a simple main effect of hemisphere \( F_{2, 55} = 7.87, p = 0.0045 \) (Fig. 6B). The sole main effect of hemisphere indicates that the N1 in frontal sources was larger (i.e., more negative) in LH across the board. Laterality of the IFG N1 revealed that clean speech was marginally lateralized in the left hemisphere \( t_{11} = -1.68, p = 0.06 \) and significantly left lateralized with moderate noise (+10 dB SNR: \( t_{11} = -2.27, p = 0.02 \)). Responses became bilateral for the +5 dB SNR condition \( t_{11} = -1.07, p = 0.85 \).
3.6. Brain-behavior correspondences

To elucidate the neuroanatomical regions that drive the observed right ear advantage in SIN perception (Fig. 2), we assessed brain-behavior correspondences separately for each ear. Right and left ear QuickSIN scores were regressed against N1 source amplitudes extracted from the two ROIs (i.e., A1 and IFG). For each ROI, responses were collapsed across hemispheres and the two noise conditions (+10 and +5 dB SNR) and regressed separately against each ear’s behavioral QuickSIN score. The pooling of hemispheres was motivated by the fact that even monaural stimulus presentation produces bilateral auditory cortical activity (Schowniesner et al., 2007). N1 amplitudes were used in this analysis given that this was the only source component that showed a significant hemisphere × SNR interaction and was clearly visible across both ROIs.

N1 responses extracted within primary auditory cortex were not correlated with behavior for either ear [LE: \( r = 0.12, p = 0.70 \); RE: \( r = -0.15, p = 0.65 \)] (data not shown). However, we found a strong association between behavioral scores and IFG (Ins./Broca’s) source activity for right [\( r = 0.63, p = 0.02 \)] but not left ear SIN listening [\( r = 0.23, p = 0.47 \) (Fig. 7)]. Despite the smaller sample size of the dataset, bootstrap and permutation resampling tests confirmed a significant difference between right and left ear correlations (\( p < 0.0001 \); Fig. 7 insets). These findings are consistent with the notions that SIN processing is driven by (i) higher order linguistic brain areas beyond sensory cortex (Bidelman and Dexter, 2015) and (ii) auditory information relayed from the right ear (Bidelman and Bhagat, 2015).

4. Discussion

4.1. Neural correlates of SIN perception

Consistent with the pattern of noise-induced changes described in previous reports (e.g., Billings et al., 2009; Parbery-Clark et al., 2011), we found that nearly all of the cortical ERPs (i.e., P1, N1, P2) showed modulations in latency and/or amplitude with changes in speech SNR at the sensor-level (i.e., electrode data; Fig. S2). However, previous dipole source analysis suggests that the mean “center of gravity” of speech evoked ERP (i.e., auditory responses) changes dramatically with SNR (e.g., Shtyrov et al., 1999). Indeed, our data revealed significant contributions of frontal brain areas during speech processing (e.g., Figs. 4 and S3) and thus, an overlap in the contributions from “auditory” and “linguistic” neural generators. This blurring of sources in scalp potentials precludes firm interpretations between auditory ERPs (sensor data) and SIN perception reported in previous studies.

Our analysis of auditory ERPs in source space allowed us to tease apart the contributions of various neural generators engaged during noise degraded speech processing. Our findings show activity restricted to primary auditory cortex undergoes distinct changes in both its amplitude and latency characteristics with noise. Nearly all deflections of the auditory ERPs became progressively weaker and delayed in more challenging listening conditions (i.e., poorer SNRs). Our results thus extend previous ERP studies (Billings et al., 2009, 2010) by demonstrating these noise-related modulations result primarily from degraded neural representations within A1.

As noted by Billings et al. (2009), the fact that the majority of ERPs are affected by SNR—despite their distinct neural generators (Picton et al., 1999)—suggest that the observed changes in cortical speech processing might be driven by more peripheral auditory structures (e.g., subcortical processing). Indeed, previous studies examining auditory brainstem responses show decreased amplitude and increased latencies of subcortical responses when processing noise-degraded compared to clean speech signals (Anderson et al., 2011; Bidelman and Krishnan, 2010; Parbery-Clark et al., 2009). Moreover, the magnitude of noise-related changes in brainstem processing predicts listeners’ success in SIN perception including recognition (Bidelman and Bhagat, 2015; Parbery-Clark et al., 2009) and discrimination tasks (Bidelman and Krishnan, 2010). In the current study, we found parallel degradations in auditory cortical source activity with additive noise. It is conceivable that this type of degraded signal analysis in early auditory cortex is at least partially inherited or influenced by structures much lower in the auditory pathway including the brainstem (Bidelman and Alain, 2015; Bidelman and Krishnan, 2010; Song et al., 2011) and cochlea (Bidelman and Bhagat, 2015).

In the current study, we found an increase in N1 negativity with low-level noise. Stronger N1 responses with decreasing SNR may reflect a partial facilitatory enhancement of speech with increasing noise levels, i.e., a physiological release from masking (Androulidakis and Jones, 2006). Consistent with our observations, low-intensity background noise has been shown in some studies to magnify cortical responses elicited by non-speech auditory stimuli (Alain et al., 2009, 2012, 2014). N1 response facilitation could emerge via efferent-induced modulations (Alain et al., 2009; Bidelman and Bhagat, 2015; Winslow et al., 1987). Indeed, medial olivococlear (MOC) efferent neurons from the
lower brainstem are thought to improve hearing in adverse listening conditions by playing an “antimasking” role in cochlear processing to improve signal detection and sensitivity in noise (Guinan, 2006; Scharf et al., 1997). The larger ERP amplitudes we observed with even minimal noise (e.g., +10 dB SNR) may reflect enhanced neural representations, vis-à-vis the efferent system, output from these lower level structures (i.e., cochlea or caudal brainstem nuclei) which are further tuned or at least maintained in cortical responses upstream. Indeed, as indexed by otoacoustic emissions, we have recently shown that SIN listening skills are well predicted by MOC efferent control (i.e., “antimasking”) to the right cochlea (Bidelman and Bhagat, 2015) suggesting that this low-level system plays an important role in degraded speech perception.

Alternatively, noise may elongate the temporal integration window for auditory information (e.g., Alain et al., 2012) which could also account for more robust ERP responses in the noise conditions. Changes in the ERPs across SNRs could also be related to other stimulus factors. For instance, changes in inter-trial phase coherence with increasing noise levels (Ponjavic-Conte et al., 2013) would affect the degree of neural synchrony and hence amplitude of speech-evoked responses. These alternate interpretations could also account for the observed changes in ERP waves for noisier speech signals.

4.2. Brain-behavior relations underlying SIN perception

Previous studies examining auditory ERPs at a sensor level have shown a relationship between the N1 component (latency) and superior SIN perception (Billings et al., 2013; Parbery-Clark et al., 2009). However as noted earlier, the diffuse, volume-conducted nature of scalp-recorded potentials makes it impossible to ascribe a single neural generator to a particular ERP deflection. Indeed, the N1 is composed of multiple sub-components (Näättänen and Picton, 1987) including sources from the frontal lobes (Picton et al., 1999). Consequently, electrode-level recordings obscure potential relations between pure auditory neural activity and SIN behavior as they contain other, non-lemniscal and non-auditory brain activity. Frontal contributions are also evident in the current data. At both the electrode- and source levels, we observed significant contribution of frontal brain regions to the auditory ERPs (e.g., Figs. 4 and S3).

The current study replicates but further extends these previous studies by isolating region-specific cortical activity and revealing the nature of these brain-behavior relations. Notably, we found that (i) inferior frontal brain areas (but not primary auditory regions) predicted behavioral SIN skills and (ii) IFG (Ins/Broca) activation was associated with QuickSIN scores in the right but not left ear. These findings support the notion that the quality of neural representation(s) and speech information routed to the left hemisphere is critical for degraded speech perception and may account for the right ear dominance in SIN listening observed in this and previous studies (Bidelman and Bhagat, 2015). Our results suggest that the correspondence between N1 and SIN performance noted in previous work (e.g., Billings et al., 2013; Parbery-Clark et al., 2009) might not be driven entirely by auditory neural representations, per se. Instead, our data lead us to infer that the degree to which linguistic brain areas continue to be engaged, dictate how forward search from higher-order information to lower-level inputs in search of representations with more optimal signal-to-noise ratio. Under the RHT, poorer neural representations that arise with noise in frontal, linguistic brain regions would force a backward search toward the sensory input (i.e., auditory cortex) in order to arrive at more favorable speech representations. While successful speech analysis may ultimately be determined by left IFG engagement (e.g., Fig. 7), the stronger resilience of left auditory cortex to noise (Fig. 5B) implies that the brain may rely more heavily on early sensory representations in LH to compensate for impoverished speech. This interpretation is supported by work showing that poorer perceivers of nonnative speech contrasts have attenuated auditory mismatch responses in frontal cortex relative to good perceivers (Bidelman and Dexter, 2015; Diaz et al., 2008) and the increased engagement of superior temporal brain regions in monolingual speakers during degraded listening conditions (Bidelman and Dexter, 2015).

4.4. Hemispheric laterality during SIN: recruitment of right- or loss of left-hemisphere involvement?

In addition to intracerebral frontal–temporal changes within each hemisphere, we found that noise degraded speech produced stronger responses in RH compared to LH (Figs. 4 and 5C). This finding is broadly consistent with previous reports demonstrating global, interhemispheric changes in speech activation during SIN listening and primacy face, that RH is progressively recruited to aid degraded speech perception in noisy listening environments (e.g., Okamoto et al., 2007; Shtryov et al., 1998, 1999). However, stronger RH activity could indicate recruitment of compensatory processing to boost SIN understanding (e.g., Du et al., 2014; Shtryov et al., 1999; Wong et al., 2009) or alternatively, the loss of linguistic function and hence residual engagement of LH resources that are more specialized to process non-speech sounds (e.g., Rinne et al., 1999). That is, a rightward shift may reflect a progressive diminishment of LH contribution, a consequence of its higher susceptibility to fluctuations in the noise envelope (Hiraiwa et al., 2008) and/or the fact that stimuli become progressively more non-speech like at lower SNRs (cf. Rinne et al., 1999).

In this regard, separating the contributions of auditory and frontal ROIs helps tease apart these two interpretations. Laterality analysis of primary auditory source activity indicated a progressive shift in neural activity from left to right hemisphere with increasing noise. Our findings thus clarify previous studies that have implied increased RH recruitment for degraded speech signals (e.g., Shtryov et al., 1998, 1999) by demonstrating this modal shift is due to changes in processing within left/right primary auditory regions. Indeed, we found that the normal leftward asymmetry in frontal sources was merely lost at poorer SNRs (becoming bilateral symmetric), whereas auditory cortical sources continued toward right lateralization with increasing noise. The asymmetry of auditory-specific neural responses is also evident in the differential change in source N1, which showed less reduction with noise in the left compared to the right auditory cortex (Fig. 5B). Collectively, our results suggest that during degraded speech processing, LH frontal (i.e., linguistic) brain areas disengage with increasing noise, relying more heavily on left auditory cortex to analyze impoverished speech signals.

Alternatively, we cannot rule out the interpretation that increased loudness of the babble masker might be more taxing in terms of neural processing, e.g., decreasing the auditory system’s ability to spectrally glimpse speech (Cooke, 2006). Spectral glimpsing might draw on right-hemispheric auditory resources more than on left-hemispheric auditory resources (Scott et al., 2009). In this interpretation, the observed hemispheric shift would not only depend on the quality of the neural representation of target speech in the left hemisphere, but also on the way the competing background noise itself is processed. Regardless of the underlying mechanism, our data demonstrate changes in the functional asymmetry of cortical speech processing during adverse
acoustic conditions and suggest that "cocktail party" listening skills are largely dependent on the quality of speech representations in the left cerebral hemisphere rather than compensatory recruitment of additional right hemisphere resources.

Acknowledgments

This work was supported by grants from the American Hearing Research Foundation (AHRF) and American Academy of Audiology Foundation (AAAF) awarded to G.M.B.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.09.020.

References


Supplemental Methods: Electrode-level responses

ERP electrode response analysis

For the purpose of data reduction, we collapsed the sensor data into five electrode clusters covering frontocentral areas of the scalp. The mean response of four adjacent electrodes within each cluster defined the five “super electrodes” across the head surface that included a left frontal (AF3, F5, F3, F1), right frontal (AF4, F2, F4, F6), left temporal (FC5, FC3, C5, C3), central (FCz, C1, Cz, C2), and right temporal (FC4, FC6, C4, C6) cluster. These clusters were selected as the auditory ERPs are maximal at the vertex and carry a prominent frontocentral scalp distribution (Picton et al., 1999; Woods, 1995). Within each electrode cluster, peak amplitude and latency were measured for the prominent deflections of the cortical response (P1, N1, P2, N2) as in the source space analysis. Analysis of the scalp data was restricted to the central electrode cluster near Cz where ERPs showed the most stereotyped morphology.

ERP laterality

We assessed laterality of the scalp-recorded waveforms at frontal electrode sites given the expected (dis-)engagement of linguistic brain regions with changes in speech SNR (Bidelman and Dexter, 2015). We extracted electrode recordings from two specific channels (F5 and FC6) in the left and right hemisphere (e.g., Herwig et al., 2003, see also http://www.brainm.com/software/pubs/dg/BA_10-20_ROI_Talairach/nearesteeg.htm). These channels were selected given their proximal location to Brodmann area 44/45 (Broca’s area) and its right hemisphere homologue based on the Talairach brain atlas (Talairach and Tournoux, 1988). We then compared left and right hemisphere responses using amplitude and latency measures as described above. A laterality index (%) was computed as LI = 100*(R-L)/(R+L), where R and L are the response measures from left and right hemispheres, respectively (cf. King et al., 1999). Responses eliciting symmetric cortical activity produce a value of zero; asymmetric responses yield a positive or negative value for right or left hemisphere dominance, respectively. P2-N2 amplitudes were used as the dependent measure for all laterality tests as this was the most prominent signature of the electrode-level data at frontal electrode sites (see Fig. S1).

Supplemental Results

ERPs (sensor-space) waveforms are shown in Fig. S1 for the various electrode clusters. Quantitative analysis of electrode-level recordings revealed that the amplitudes of the earlier waves including the P1 \([F_{2, 22} = 4.97, p = 0.0165]\) and N1 \([F_{2, 22} = 10.59, p = 0.0006]\) were modulated by speech SNR. Whereas P1 decreased in magnitude with increasing noise, N1 increased in magnitude, becoming more negative at lower SNRs (Fig. S2A). Neither the P2 \([F_{2, 22} = 2.38, p = 0.1163]\) nor N2 \([F_{2, 22} = 1.98, p = 0.16]\) components showed appreciable amplitude changes with SNR.
ERP latency was similarly modulated by noise level (Fig. S2B). With the exception of N2, all component waves showed noise-related changes in latency, occurring later for poorer compared to more favorable SNRs [P1: \( F_{2, 22} = 4.69, p = 0.02 \); N1: \( F_{2, 22} = 5.59, p = 0.011 \); P2: \( F_{2, 22} = 14.67, p < 0.001 \); N2: \( F_{2, 22} = 0.04, p = 0.96 \)].

Figure S1: Distribution of the speech-evoked ERPs as a function of SNR in five electrode regions of interest around the scalp. Grey insets, time waveforms of the speech stimulus. Top right, electrode cluster definition. Prominent latency and amplitude modulations are observed at the vertex (central cluster), particularly in the N1 and later deflections.

Figure S2: Amplitude (A) and latency (B) profiles for the speech-evoked ERPs with increasing noise. Prominent amplitude decrements and prolonged latencies are seen in the variance component waves at poorer speech SNRs. *\( p < 0.05 \), **\( p < 0.01 \), ***\( p < 0.001 \)
**Hemispheric laterality of scalp-recorded ERPs during SIN processing**

A comparison of speech-ERPs recorded in right vs. left hemispheres is shown in Figure S3A-C as a function of speech SNR. Traces reflect neuroelectric activity recorded from electrodes proximal to Broca’s area and its right hemisphere homologue (Herwig et al., 2003; Talairach and Tournoux, 1988). For clean speech, more robust responses are observed in the left hemisphere compared to the right (i.e., LH > RH). However, consistent with scalp topographies (i.e., Fig. 3), noise-degraded speech produced a rightward asymmetry (i.e., RH > LH) in the distribution of the scalp potentials. These findings indicate a modal shift in hemispheric dominance in frontal brain regions with increasing noise level (i.e., decreasing SNR).

![Figure S3: Hemispheric laterality for noise-degraded speech processing.](image)

**Figure S3:** Hemispheric laterality for noise-degraded speech processing. (A-C) Hemispheric comparison of cortical ERPs recorded proximal to Broca’s area (electrode F5) and its right hemisphere homologue (electrode FC6) for (A) clean and degraded speech at (B) +10 dB SNR and (C) +5 dB SNR. (D) Laterality of the cortical ERPs (P2-N2 magnitudes). A leftward laterality (i.e., LH > RH) is observed for clean speech. The functional asymmetry in auditory processing becomes more symmetric and then shifts rightward for progressively noisier speech.*p < 0.05, **p < 0.01

Functional laterality of the cortical ERPs during SIN processing is shown in Figure S3D. Note that |LI| ≥ 0.2 is typically considered “lateralized” (Seghier, 2008). Quantitative analyses confirmed a shift in hemispheric dominance with increasing levels of noise. Whereas clean speech was strongly lateralized in the left hemisphere \(t_{11} = -3.51, p = 0.0024\), the introduction of noise produced a shift toward the right hemisphere; laterality diminished (became more bilateral) at +10 dB SNR \(t_{11} = 0.21, p = 0.58\) and began showing a rightward dominance at +5 dB SNR \(t_{11} = 1.69, p = 0.05\). This was supported
by an ANOVA, which revealed a significant effect of SNR on ERP laterality [$F_{2, 22} = 12.36, p = 0.0003$]. Follow-up contrasts confirmed significant change in speech laterality with increasing noise [linear contrast of SNR: $t_{11} = 3.71, p = 0.0034$]. Together, these findings indicate that auditory cortical responses (at the scalp) are weighted more heavily in the left than right hemisphere when processing clean, unadulterated speech. Furthermore, the leftward response asymmetry becomes more bilateral and shifts progressively rightward during adverse speech listening.

References


