Title: External and Internal Signal-to-noise Ratios Alter Timing and Location of Cortical Activities During Speech-in-noise Perception

Abbreviated title: Neural correlates of speech-in-noise performance

Subong Kim,[1,#] Adam T. Schwalje,[2,#] Andrew S. Liu,[2] Phillip E. Gander,[3] Bob McMurray, [1,2,4] Timothy D. Griffiths,[5] and Inyong Choi[1,2,*]

[1] Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA 52242, USA

[2] Department of Otolaryngology – Head and Neck Surgery, University of Iowa Hospitals and Clinics, Iowa City, IA 52242, USA

[3] Department of Neurosurgery, University of Iowa Hospitals and Clinics, Iowa City, IA 52242, USA

[4] Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA 52242, USA

[5] Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

[#] SK and ATS contributed equally to this work.

[*] Corresponding author at: Department of Communication Sciences and Disorders, University of Iowa, 250 Hawkins Dr., Iowa City, IA 52242, USA. Email address: inyong-choi@uiowa.edu

Number of pages: 34

Number of figures: 6

Number of words: 238 (abstract), 1185 (introduction), and 1862 words (discussion)

Author contributions:

S.K. and A.T.S. contributed equally to this work. I.C. designed the experiments. S.K. ran the experiments. S.K., A.T.S., A.S.L., P.E.G., B.M., T.D.G., and I.C. analyzed and interpreted data.

S.K., A.T.S., and I.C. prepared the manuscript, with revisions and suggestions from A.S.L, P.E.G., B.M., and T.D.G..

Acknowledgments

This work was supported by Hearing Health Foundation Emerging Research Grant and departmental start-up awarded to Inyong Choi, and NIH T32 (5T32DC000040-24) and NIDCD P50 (DC000242 31A1) awarded to Bruce Gantz. The authors declare no competing financial interests.

Abstract

Understanding speech in noise (SiN) is a complex task that recruits multiple cortical subsystems. There is a variance in individuals' ability to understand SiN that cannot be explained by simple hearing profiles, which suggests that central factors may underlie the variance in SiN ability. Characterizing central functions that exhibit individual differences during a SiN task and finding their relative contributions to predicting SiN performance can reveal key neural mechanisms of SiN understanding. Here, we elucidated a few cortical functions involved during a SiN task and their hierarchical relationship using both within- and across-subject approaches. Through our within-subject analysis of source-localized electroencephalography, we demonstrated how acoustic signal-to-noise ratio (SNR) alters neural activities along the auditory-motor pathway, or dorsal stream, of speech perception. In quieter noise, left supramarginal gyrus (SMG, BA40) exhibited dominant activity at an early timing (~300 ms after word onset). In contrast, in louder noise, left inferior frontal gyrus (IFG, BA44) showed dominant activity at a later timing (~700 ms). Further, through an individual differences approach, we showed that listeners show different neural sensitivity to the background noise and target speech, reflected in the amplitude ratio of cortical responses to speech and noise, named as an "internal SNR." We found the "softer noise" pattern of activity in listeners with better internal SNR, who also performed better. This result implies that how well a listener "unmask" target speech from noise determines the subsequent speech analysis and SiN performance.

Significance

This study elucidated crucial cortical mechanisms underlying speech-in-noise perception using both within- and across-subject design approaches. We found that cortical auditory evoked responses to speech involved early activation in the temporo-parietal cortex in an easy condition while a hard condition cortical activity involved late activation in the frontal cortex. Importantly, the across-subject analysis showed that pre-speech time cortical activity predicts post-speech

time activity, in such a way that good performers with better neural suppression of background noise show cortical activity similar to the pattern observed in the easier condition regardless of given acoustic noise level. This suggests a critical role of pre-lexical sensory gain control processes affecting performance and cognitive load during speech-in-noise perception.

Introduction

Understanding speech in noise (SiN) is essential for communication in social settings. Yet young normal-hearing listeners are remarkably good at this: even in challenging SiN conditions where the speech and noise have the same intensity (i.e., 0 dB signal-to-noise ratio: SNR) and overlapped frequency components, they often recognize nearly 90% of sentences correctly (Ohlenforst et al., 2017). This suggests a remarkable capacity of the auditory system to cope with noise. However, the ability to understand SiN degrades severely with increased background noise level (Ohlenforst et al., 2017), hearing loss (Harris and Swenson, 1990), and/or aging (Nabelek, 1988). Moreover, recent studies show that normal hearing listeners show large individual differences in SiN performance (Liberman et al., 2016). By linking this variable ability for SiN perception to cortical activity, we may be able to understand the neural mechanisms by which humans accomplish this ability, and this may shape our understanding of how best to remediate hearing loss.

Different neuro-cognitive mechanisms might give rise to better or worse SiN performance. First, listeners may vary in the mechanisms for representing sound in the ascending pathway to the auditory cortex (AC) and in auditory scene analysis (Bregman, 1999), both of which are required to separate signal from noise. Auditory scene analysis processes can inhibit neural responses to task-irrelevant sensory inputs even before the target sound is heard, based on expectations of differences between the target and masker such as differences in their spectra (Lee et al., 2013), location (Frey et al., 2014; Goldberg et al., 2014), and timing (Lange, 2009). A successful auditory scene analysis during a SiN task will unmask the target speech from maskers, which will enhance effective signal-to-noise ratio (SNR) in the neural pathway. Second, listeners might vary in neural mechanisms for representing speech in the temporo-parietal-frontal language network that might compensate for a noisy signal. Current models of the neural processing of speech suggest two distinct cortical networks (i.e., dorsal and ventral stream) for speech processing (Scott and Johnsrude, 2003; Hickok and Poeppel, 2007; Myers

et al., 2009; Gow, 2012). Such a division of labor is highlighted by work showing that speech in noise engages the dorsal stream more strongly than in quiet (Hickok and Poeppel, 2007; Liebenthal et al., 2013; Bidelman and Howell, 2016; Du et al., 2016). However, we have limited knowledge of the functional roles of these two cortical pathways, their timing, and their hierarchical relation.

Signal separation and compensatory mechanisms can be distinguished by using functional neuroimaging in an individual differences approach. That is, we can compare the degree to which accuracy in a SiN task is correlated to either pattern of activity in auditory and related (signal analysis), or with frontal areas involving articulation and decision making (compensation)[1]. A few have studies asked how individual differences in cortical pathways correlate with SiN performance. These suggest activity in frontal areas (e.g., inferior frontal gyrus: IFG) predict SiNs performance (Wong et al., 2009; Bidelman and Howell, 2016; Du et al., 2016). They largely did not find an influence of lower-level areas; however, as we discuss, methodological limits may have prevented this. Importantly, no work has examined both dorsal and ventral processes simultaneously to determine if signal separation or compensatory speech pathways capture unique variance in performance. Our central hypothesis is that the quality of early signal analysis that occurs before auditory-motor transformations, rather than variation in later compensatory processes dependent on the dorsal pathway, uniquely predicts speech perception accuracy.

A secondary test of the importance of each pathway is the relative timing of activity in these pathways during speech processing. Functionally, work using eye-movements in the

[1] A third possible mechanism—differences in auditory attention—likely spans both networks. Auditory attention likely originates in a fronto-parietal network involving the inferior frontal gyrus (IFG), the superior parietal sulcus, and the intraparietal sulcus (Teki et al., 2011; Hausfeld et al., 2018), but affects early-stage auditory activity (Choi et al., 2013; Choi et al., 2014; Bressler et al., 2017). Under an attentional account, if the most variance amongst listeners is due to differences in deploying attention, we should see SiN performance primarily correlated with frontal activity; whereas if most variance is due to using attention to clean up the signal it should primarily be associated with the auditory and related cortex.

visual world paradigm has extensively characterized the time course of word recognition in both quiet (Allopenna et al., 1998; Dahan and Gareth Gaskell, 2007; Magnuson et al., 2007) and under challenging conditions such as noise or signal degradation (Huettig and Altmann, 2005; Ben-David et al., 2011; McQueen and Huettig, 2012; Brouwer and Bradlow, 2016; McMurray et al., 2017). In general, this work suggests that in quiet, listeners activate a range of lexical candidates immediately at the onset of the auditory stimulus. This *lexical competition* resolves rapidly by around 250 ms after the uniqueness point of the word. A moderate amount of degradation or noise typically imposes about a 75-100 ms delay on this recognition process (Ben-David et al., 2011; Farris-Trimble et al., 2014), where severe degradation can delay lexical access by up to 250 ms (Farris-Trimble et al., 2014; McMurray et al., 2017). This is particularly relevant for evaluating the causal role of downstream compensatory processes – if such processes are later than 400-500 ms, this may be too late to reflect lexical competition. However, the timing of cortical activity within each pathway is largely unknown, as most of the work on speech in noise perception has been conducted with functional magnetic resonance imaging (fMRI) (Wong et al., 2008; Wong et al., 2009; Du et al., 2014, 2016) which has the poor temporal resolution.

Assuming that left supramarginal gyrus (SMG: the anterior part of the inferior parietal lobule) is an early stage in the dorsal pathway works as an interface between auditory/phonological representations in the superior temporal gyrus and motor ones in the frontal lobe (Binder et al., 2004; Gow, 2012), we expect to see early SMG activity in the less adverse listening situation. In contrast, we hypothesize that noisier listening conditions will evoke late activities in the frontal area, reflecting downstream compensatory processes.

The present study tests above hypotheses using both a within-subject design and individual differences approaches. First, we identify how both primary auditory pathways and frontal compensatory processes differ in noise. This within-subject design examined the effect of acoustic SNR on the a) timing and b) location of cortical activity during word-in-noise

recognition. For this, we used two distinct high-density electroencephalographic (EEG) analyses: 1) Hypothesis-driven source estimation that examined time courses of evoked responses within two regions-of interest (ROIs) – SMG and IFG; and 2) sensor-space microstate analysis as a data-driven approach that cross-validates the ROI-based analysis. Next, we assessed the relative and unique contributions of both primary and compensatory cortical processes in predicting SiN performance. We use the same EEG data to quantify an individual's speech unmasking ability by computing the ratio of cortical evoked responses to noise- and target speech-onset, or "internal SNR." We also quantify post-speech-time neural activity in the dorsal speech-motor pathway. These were then used in a regression analysis to determine the relative contribution of each to SiN performance.

Materials and Methods

Participants

All study procedures were reviewed and approved by the local Institutional Review Board. Thirty subjects between 19 and 31 years of age (mean = 22.4 years, SD = 2.8 years; median = 22 years; 9 (30%) male) were recruited from a population of students at a large Midwestern university. All subjects were native speakers of American English, with normal hearing thresholds no worse than 20 dB HL at any frequency, tested in octaves from 250 to 8000 Hz. Four subjects (1 male) were excluded from the analysis due to signal contamination across several EEG channels.

Task design and procedures

We simultaneously measured behavioral performance and cortical neural activity a short (15 minute) experimental sessions.

Each trial (**Figure 1**) began with the presentation of a fixation cross ('+') on the screen. Listeners were asked to fix their gaze on this throughout the trial to minimize eye-movement

artifacts. Next, they heard the cue phrase “check the word.” This enabled listeners to predict the timing of next acoustic event (the noise onset). After fixed-duration (700 ms) silence that followed the cue phrase, multi-talker babble noise started and continued for 2 seconds. One second after the noise onset, the target word was heard 100 ms after the composite auditory stimulus (noise + word) offset, four written choices appeared on the screen. The response options differed either in the initial or the final consonant (e.g., ‘than,’ ‘van,’ ‘ban,’ and ‘pan,’ for target word ‘ban’; ‘hit,’ ‘hip,’ ‘hiss,’ ‘hitch’ for target word ‘hiss’). Subjects pressed a button on a keypad to indicate their choice. No feedback was given to the subject at the end of a trial. The next trial began 1 second after the button press. This trial structure was intended to minimize visual, pre-motor, and motor artifacts during the time of interest surrounding the auditory stimuli. The timing and intervals of auditory stimuli (i.e., cue phrase, noise, and target) were intended to derive well-distinct cortical evoked responses to the onsets of background noise and target word.

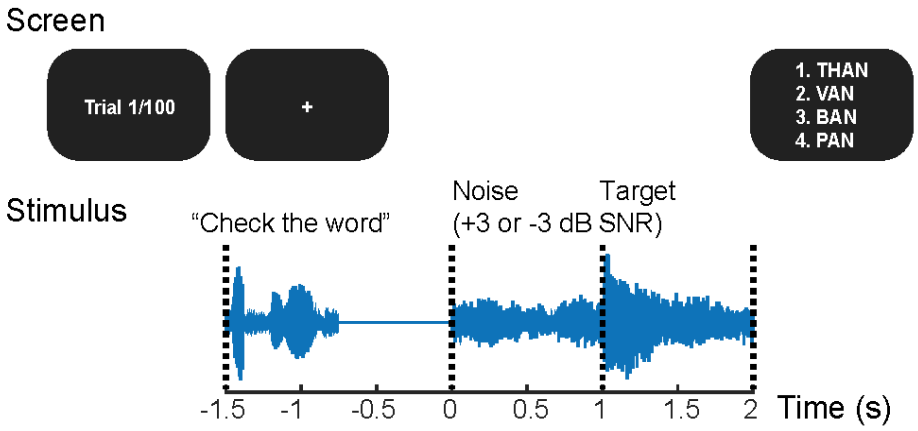


Figure 1. Trial and stimulus structure. Every trial starts with the cue phrase “check the word.” A target word starts 1 second after the noise onset. Four choices are given after the word ends; subjects select the correct answer with a keypad. No feedback is given. The noise level is manipulated to create high (+3 dB) and low (-3 dB) SNR conditions. Subjects complete 50 trials for each condition.

Since we are particularly interested in SMG/IFG regions that are involved in phonological/lexical processing (Hickok and Poeppel, 2007), we elected to use natural monosyllabic words, rather than simpler non-sense speech tokens used by prior studies

(Parbery-Clark et al., 2009; Bidelman and Howell, 2016). This engaged lexical processing, placed high demands on cortical processing, and maximized ecological validity (Gagne et al., 2017). Thus, target words consisted of 100 hundred CVC words from the California Consonant Test (CCT) (Owens and Schubert, 1977), spoken by a male speaker with a General American accent.

The RMS level of noise was either 68 and 62 dB SPL, and target word was always presented at 65 dB SPL. This led to either +3 or -3dB SNR (referred to as "high SNR" and "low SNR," respectively). Fifty words were presented at each SNR (±3 dB). -3 dB SNR was chosen to emulate a highly effortful listening condition yielding ~70% accuracy from pilot experiments. +3 dB SNR condition emulates an easy listening condition.

The task was implemented using the Psychtoolbox 3 package (Brainard, 1997; Pelli, 1997) for Matlab (R2016b, The Mathworks). Participants were tested a sound-treated, electrically shielded booth with a single loudspeaker (model #LOFT40, JBL) positioned at a 0° azimuth angle at a distance of 1.2 m. A computer monitor was located 0.5m in front of the subject at eye level. The auditory stimuli were presented at the same levels for all subjects.

EEG acquisition and preprocessing

Scalp electrical activity (EEG) was recorded during the SiN task using the BioSemi ActiveTwo system at a 2048 Hz sampling rate. Sixty-four active electrodes were placed according to the international 10-20 configuration. Trigger signals were sent from Matlab (R2016b, The Mathworks) to the ActiView acquisition software (BioSemi). The recorded EEG data from each channel were bandpass filtered from 1 to 50 Hz using a 2048-point FIR filter. Epochs were extracted from -500 ms to 3 s relative to stimulus onset. After baseline correction using the average voltage between -200 and 0 ms, epochs were down-sampled to 256 Hz.

Since we were interested in the speech-evoked responses from frontal brain regions, we opted for a non-modifying approach to eye blink rejection: Trials that were contaminated by an

eye blink artifact were rejected based on the voltage value of the Fp1 electrode (bandpass filtered between 1 and 20 Hz). Rejection thresholds for eye blink artifacts were chosen individually for each subject, and separately for the noise and target word periods. After rejecting bad trials, grand averages for each electrode were calculated for the two conditions. For analysis of speech-evoked responses, we repeated baseline correction using the average signal in the 300 ms preceding the word onset.

Source analysis

The source-space analysis was based on minimum norm estimation (Gramfort et al., 2013; Gramfort et al., 2014) as a form of multiple sparse priors (Friston et al., 2008). After co-registration of average electrode positions to the reconstructed average head model MRI, the forward solution (a linear operator that transforms source-space signals to sensor space) was computed using a single-compartment boundary-element model (Hämäläinen, 1989). The cortical current distribution was estimated assuming that the orientation of the source is perpendicular to the cortical mesh. Cross-channel EEG-noise covariance, computed for each subject, was used to calculate the inverse operators. A noise-normalization procedure was used to obtain dynamic statistical parametric maps (dSPMs) as *z*-scores (Dale et al., 2000). The inverse solution estimated the source-space time courses of event-related activity at each of 10,242 cortical voxels per hemisphere. In the present study, two predetermined ROIs will be included: (1) bilateral SMG, and (2) bilateral pars opercularis and pars triangularis of IFG. The SMG is an early stage in the dorsal pathway and, among its many roles, works as an interface between auditory/phonological representations in the superior temporal gyrus and motor ones in frontal lobe including precentral/postcentral and IFG (Binder et al., 2004; Gow, 2012). To ensure the fidelity of source localization at our ROIs, we applied the same analysis to HG, which expected to be active in auditory tasks, before running ROI-based source analysis explained below (**supplement Figure 1**).

234

235 ***Statistical analyses***

236 ***ROI-based source analysis.*** For ROI-based source analysis, one-sample *t*-tests were
237 used to compare the distributions of mean source event-related potential (ERP) to zero
238 (baseline voltage) in each SNR condition. The t-value envelope was then computed (as a form
239 of smoothing). This was done by applying a bandpass filter, then calculating the absolute value
240 of the Hilbert transform. The bandpass filter was set to one of two center frequencies,
241 depending on the specific ROIs. Because the neural oscillations evoked by early ERP
242 components such as N1-P2 have peak latencies of about 100 ms at their half cycle
243 (approximately 5 Hz), and late ERP components such as N2-P3 have latencies of about 200 ms
244 at their half cycle (approximately 2.5 Hz), the bandpass filter was set to between either 3 to 7 or
245 1 to 5 Hz to highlight these components. The *t*-value envelope calculated using the bandpass
246 filter between 3 to 7 Hz was used to investigate HG and SMG ROIs, and earlier times of
247 interest, while the envelope using the 1 – 5 Hz filter was used for IFG ROI and later times of
248 interest. For each SNR condition, the whole brain *t*-value envelope time courses were obtained.
249 Since we did not have individual structural MRI head models, it was not ideal to take the
250 summed activity (mean or median) for all the voxels within ROIs. This is because individual
251 difference in functional and anatomical structure of the brain may result in spatial blurring since
252 current densities across adjacent voxels can overlap each other. Instead, representative voxels
253 were identified for each ROI, for each SNR condition. We used a combination of previously-
254 described methods to select voxels of interest that were used in fMRI studies (Tong et al.,
255 2016). The cross-correlation coefficients for *t*-value envelopes between all voxels in an ROI
256 were calculated across time (up to 800 ms after the target word onset), and then the mean
257 coefficient was calculated for each voxel. The most representative voxel was defined as having
258 the maximum value mean coefficient, while also being above threshold at two or more

259 continuous timepoints based on voxel's p value, as determined using one-sample *t*-tests (Tong
260 et al., 2016).
261 Once the most representative voxel was chosen for each SNR condition, a leave-one-
262 out procedure (i.e., Jackknife approach) was used to compare the population means between
263 the two SNR conditions, at those voxels, using paired t-tests. Prior to computing *p*-values, t-
264 statistics were adjusted for jackknifing (Luck, 2014). The positive false discovery rate (pFDR)
265 was estimated from those p-values after downsampling the time sequence according to Nyquist
266 theorem, and used to find timepoints that showed significant difference between SNR conditions
267 (Storey, 2002). Finally, the whole cortical surface source space was evaluated at those
268 timepoints.
269
270 ***Microstate analyses.*** Microstate analysis was conducted to cross-validate our ROI-
271 based analyses that assess effect of external SNR on ERPs at the group level. Microstates
272 describe transient, quasi-stable topographic orientations which provide information about the
273 timing of cognitive processes (Koenig et al., 2014). Microstates have been used to characterize
274 both resting state and event-related EEG activity (Ott et al., 2011; Schiller et al., 2016). The
275 microstate analysis was conducted on grand mean data (averaged across all subjects),
276 separately in each SNR condition and were implemented in the RAGU program (Koenig et al.,
277 2011; Koenig et al., 2014).
278 To conduct this analysis, we first identified four microstate cluster maps based on spatial
279 clustering of ERP topographies. To this, a k-means algorithm was used for microstate
280 identification with ten random re-initializations. As a cross-validation procedure, the 26 subjects
281 were randomly split 50 times into a training set and a test set, each comprising 13 subjects.
282 Next, the grand mean voltages at each timepoint of each SNR condition were assigned to the
283 best fit cluster map (Koenig et al., 2011; Koenig et al., 2014). To do this, each timepoint of the

ERP was assigned to the specific microstate cluster map that had the highest correlation coefficients with the topography of the ERP, across all electrodes, at that timepoint.

To assess how well the grand mean ERPs were explained by the different microstate clusters, we calculated the relative area of global field power (GFP) for each cluster, after assigning the timepoints to microstates. The GFP at time *t* is equal to the standard deviation of the signal at all *N* electrodes, which is defined as

$$(1)\quad \mathrm{GFP}(t) = \sqrt{\frac{\sum_{i=1}^{N}(v_i(t)-\bar{v}(t))^2}{N}},$$

where $v_i(t)$ is the voltage at electrode *i*, and $\bar{v}(t)$ is the average voltage across all electrodes at time *t*.

After identification of the microstate explaining the most variance for grand average ERP in each SNR condition, the timepoints of maximum GFP for that microstate were used to create whole brain maps showing cortical source activity. The *t*-value envelope calculated by one-sample t-tests in each SNR condition was used to investigate the source activity at early timepoints (about up to 400 ms after the word onset) using the bandpass filter between 3 to 7 Hz, and the source activity at later timepoints (about after 400 ms) using the 1 – 5 Hz filter.

Regression approaches. To conduct multiple linear regressions, we used a jackknifing approach (Stahl and Gibbons, 2004; Luck, 2014). In this approach the relevant neural factors were computed for all subjects but one. This was repeated leaving out each subject in turn, and the resulting data submitted to a linear regression with SiN performance (accuracy) as the dependent variable, and SMG/IFG activation, and internal SNR as the predictor variables. Test statistics were then adjusted to account for the fact that each data-point represents N-1 participants.

Correlation/regression analyses used these techniques to simultaneously examine bottom-up and compensatory related SiN performance to three factors. Our first, analysis

examined raw cortical evoked activity to the target and noise. To best represent bilateral auditory cortical activity, we used sensor-space ERP envelopes using the 3 – 7 Hz bandpass filter from channel Cz at around 200 ms after the noise onset and at about 200 ms after the target word onset, based on the timing determined by microstate analysis. Then, "internal SNR" was defined as the ratio of target word-evoked ERP envelope peaks to noise-evoked ERP envelope peaks magnitude described above, in dB scale, obtained from channel Cz (Equation 3). The internal SNR is different for each subject, and is separate from the fixed external, or acoustic, SNR (here, ±3 dB).

$$\text{(2)}\quad Internal\ SNR = 20log_{10}\frac{\text{Word evoked potential}}{\text{Noise evoked potential}}$$

Second, to examine dorsal regions, we used cortical regions that were previously identified in the ROI-based source analysis described above. The peak magnitudes of the *t*-value envelopes were obtained for early SMG activation in the high SNR condition and late IFG activation in the low SNR condition.

Results

SiN performance

There was a large variance in performance among participants. This was observed in both the high SNR condition (mean accuracy = 80.64%, SD = 7.81%, median = 83.01%; mean reaction time = 1.53 s, SD = 0.32 s, median = 1.55 s) and the low SNR condition (mean accuracy = 68.21%, SD = 8.92%, median = 70.37%; mean reaction time = 1.70 s, SD = 0.36 s, median = 1.69 s). Both accuracy ($t(25) = 6.99$, $p < 0.001$, paired t-test) and reaction time ($t(25) = -3.81$, $p < 0.001$, paired t-test) differed significantly between the two SNR conditions (**Figure 2A**). Reaction time and accuracy were correlated in the high SNR condition (**Figure 2B**, Pearson correlation $r = -0.50$, $p = 0.009$), but not in the low SNR condition (**Figure 2C**, $r = -0.19$,

p = 0.34). As a whole, these results validate that the SNR manipulation was sufficient to create more challenging speech perception conditions.

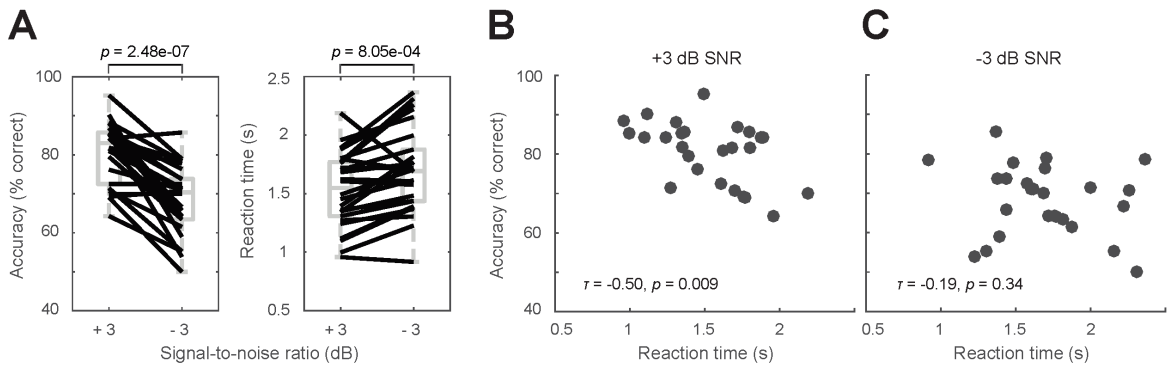


Figure 2. Behavioral results. **A.** Summary of behavioral performance for the two conditions (+3 and -3 dB SNR). Boxes denote the 25th – 75th percentile range; the horizontal bars in the center denote the median; the ranges are indicated by vertical dashed lines. Solid lines connect points for the same subject in different conditions. **B.** Average accuracy as a function of reaction time in +3 dB SNR condition. **C.** Average accuracy and reaction time in -3 dB SNR condition.

The effect of SNR on cortical activity

To assess the cortical activity, we converted sensor-space EEG signals to whole brain source-activity. This allowed us to localize the effects of noise on targeted ROIs at specific times. Within left hemisphere SMG, the high SNR condition showed significantly greater activity than the low SNR condition from 200 to 330 ms (FDR adjusted p < 0.05) (**Figure 3B left**). Within left hemisphere IFG, the low SNR condition showed significantly greater activity than the high SNR condition from 740 to 830 ms (FDR adjusted p < 0.05) (**Figure 3B right**), throughout triangular and opercula regions.

A visual representation of this finding can be seen in **Figure 3C**. Here, *t*-values reveal significant differences between high and low SNR conditions. The timepoints were chosen for display where the greatest number of voxels have significant differences. We observed increased amplitude of the left hemisphere SMG activation in the high SNR condition at around 250 ms (*t*-value envelope). This may indicate efficient lexical processing in a relatively favorable listening condition. However, in the low SNR condition, the peak amplitude of the left hemisphere IFG activation (*t*-value envelope) seen around 700 ms after word onset. Given most target words were around 500 ms in duration, SMG was primarily activated during the presentation of the target word, while IFG was activated after its offset.

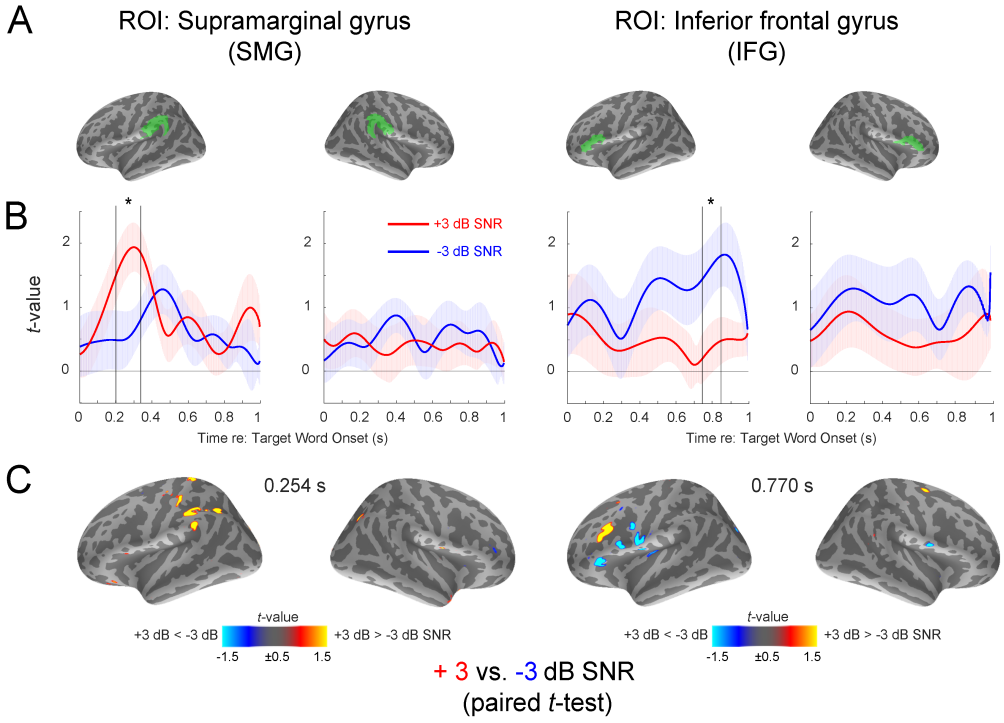


Figure 3. Region-of-interest (ROI) based source analysis. **A.** Cortical labels for two ROIs in left and right hemispheres: supramarginal gyrus (SMG), and the pars opercularis and triangularis of the inferior frontal gyrus (IFG), respectively. **B.** The time course of the *t*-value envelope, with the standard error of the mean (±1 SEM), obtained at representative voxels in each SNR condition (red color: +3 dB SNR, blue color: -3 dB SNR). An asterisk shows the timing of significant difference between +3 and -3 dB SNR conditions (paired *t*-test, FDR adjusted $p < 0.05$). **C.** Whole brain maps showing statistical contrasts (*t*-values obtained from paired *t*-tests between the two SNR conditions) of source activation at each voxel, only displaying those with *p*-value less than 0.05, at the timepoint that shows significant differences over the broadest area in the ROIs within the time range described above.

To more precisely timelock these neural events to the stimulus, the webMAUS (Kisler et al., 2017) was used to identify the starting location of second and third phonemes in each of the 100 stimuli. A histogram of these acoustic time points is shown in **Figure 4A**. **Figure 4B** shows the timing of SMG and IFG activity relative to the distribution of phoneme onsets in the target word stimuli. This confirms that the early SMG activation occurs within the timecourse of target words before the final phoneme is presented; the late IFG activation occurs after all words are presented.

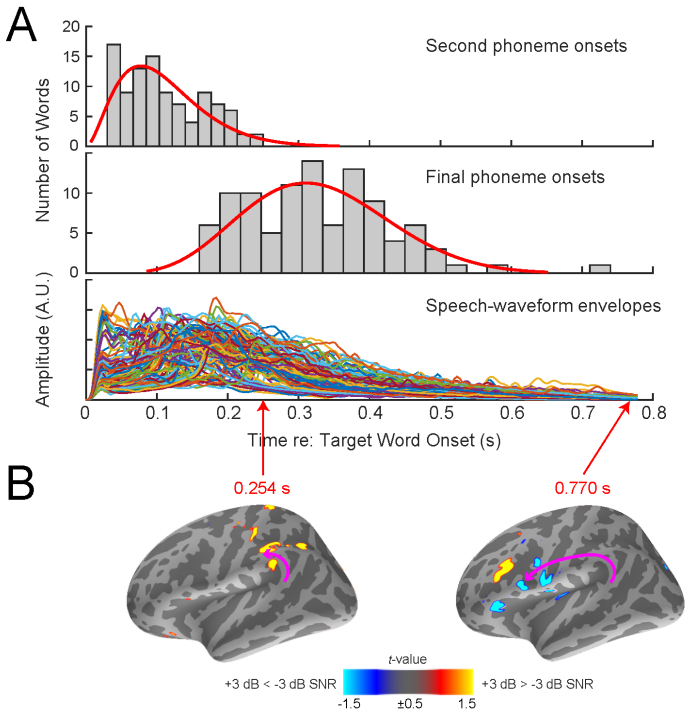


Figure 4. Timings of significant cortical activity relative to distributions of phonological events. **A.** Top and second panel show a histogram of the onsets of second and final phoneme of each stimulus. The third panel shows superimposed temporal envelopes extracted from waveforms of the 100 words. **B.** The whole brain maps at the bottom are from **Figure 3C** that shows statistical contrasts of source activation at the timepoints that show significant differences between the two SNR conditions. Purple curves on the cortical maps represent the conceptual illustration of ascending information flow through the dorsal pathway.

External (acoustic) SNR effects on timepoints and regions of interest based on microstate analysis

To further assess temporal dynamics of neural activity during SiN perception in a data-driven way, and to cross-validate our ROI-based analyses, we performed microstate analysis (Lehmann, 1989b, a; Wackermann et al., 1993). Microstate analysis clusters time series of ERP data into multiple different brief brain states, which may indicate different stages of the information processing (Lehmann, 1989b). Four microstate maps were identified. (**Figure 5C**). The grand mean ERP at each timepoint was assigned to one of the microstate clusters, and GFP was calculated at those timepoints (**Figure 5B, C**).

Calculation of the relative area of GFP revealed that microstates 1 and 2 explained the largest variance in sensor-space ERPs over time in low SNR (area = 37%) and high SNR (area = 37%) condition, respectively. The timing, suggested by maximum GFP and the maximum deflection of ERPs at frontal-central electrodes among the timepoints assigned to microstate 1, was 0.668 seconds after word onset for the low SNR condition. The timing for microstate 2 was 0.320 seconds after word onset for the high SNR condition.

We next conducted a whole-brain source analysis at the timepoint assigned to microstate 1 using one-sample t-tests against 0. This revealed increased activity in left hemisphere IFG activation for the low SNR condition (**Figure 5D**). The same analysis at the timepoint assigned to microstate 2 showed increased activity in the left hemisphere SMG for the high SNR condition (**Figure 5D**). These results from microstate analysis are consistent with the results of ROI-based analysis (illustrated in **Figure 3**) regarding timings and regions of neural activation.

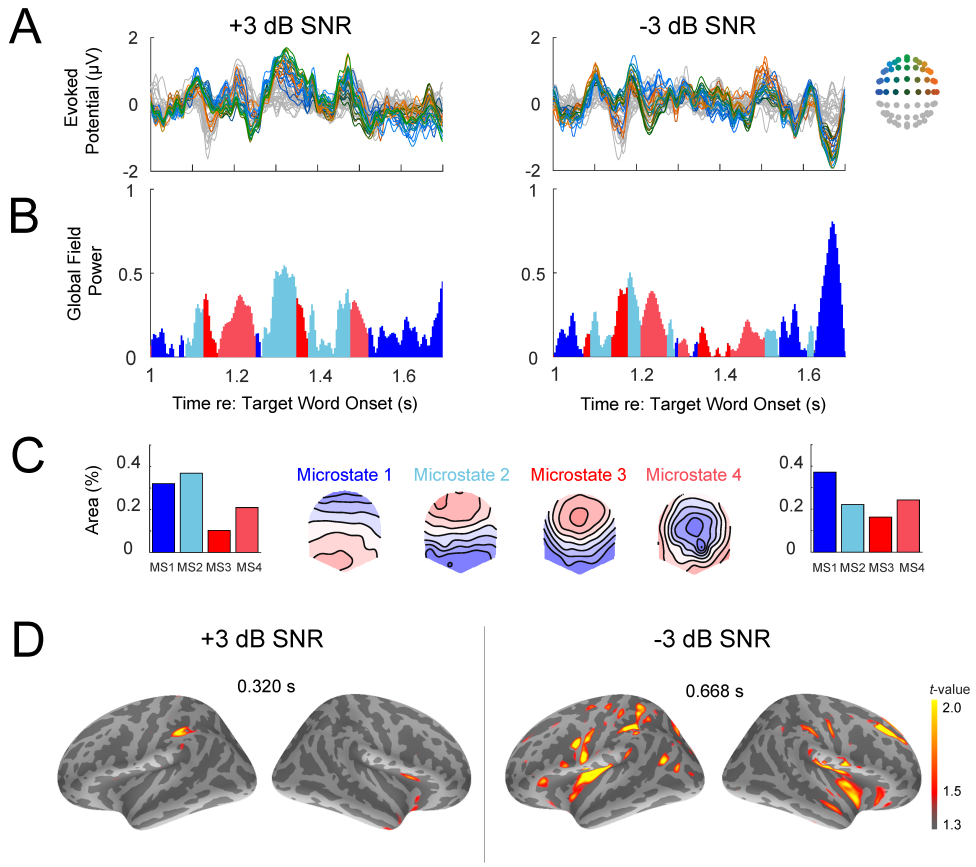


Figure 5. Microstate analysis. **A.** Evoked responses over time after the word onset for +3 and -3 dB SNR condition. Each color represents ERPs from a different channel of interest. **B.** Global field power (GFP) is calculated at each timepoint that is assigned to one of the microstate clusters. **C.** Four microstate cluster maps. Dark blue, light blue, red, and dark pink colors represent microstates 1, 2, 3, and 4, respectively. The relative area of GFP is calculated and reveals the highest value for the microstate 1 and 2 for -3 and +3 dB SNR condition, respectively. **D.** Whole brain maps obtained at the times assigned to microstates 1 and 2, that show maximum GFP and the maximum peak of ERPs at the frontal-central electrodes.

383

384 ***Individual differences in internal SNR predict SiN performance***

385 For visualization purposes, we identified good and poor performers by conducting a
386 median split based on their performance in the task. **Figure 6A** shows GFP of the grand mean
387 ERPs for good and poor performance and topographies obtained at two timepoints identified by
388 microstate maps in the low SNR condition: 1) 220 ms after the noise onset, corresponding to the
389 auditory P2 to the noise; and 2) 240 ms after target word onset, corresponding to the AC-driven

N1 response. The word-evoked N1 was chosen as the first clearly AC-driven response to the word onset as evidenced by its assignment to microstate 4 in the previous analysis, while the timing of the noise-evoked P2 was suggested by microstate 3. Despite the same noise level for these two groups of subjects, good performers exhibited less AC response to the background noise, and greater AC response to the target word at central channels including Cz. This validates that each component of the internal SNR measure seems to contribute separately to SiN performance.

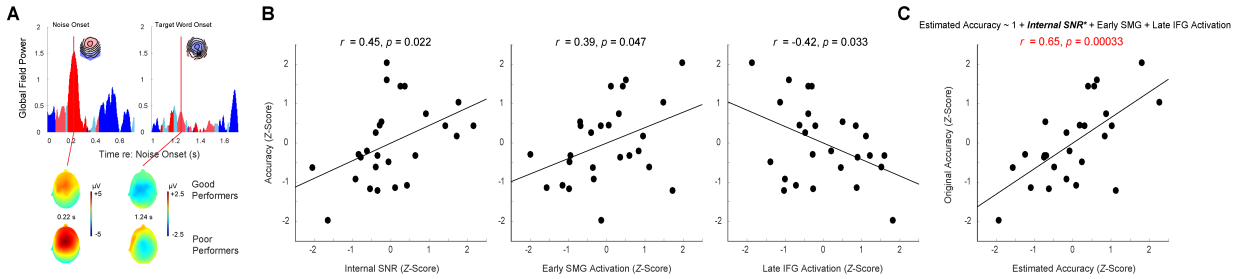


Figure 6. Individual differences in speech-in-noise processing. **A.** Global field power of the grand mean evoked potentials after the noise onset and after the target word onset, separately in the low SNR condition. Scalp topographies were examined at the timepoints, suggested by microstate analysis from **Figure 5**, and compared between good and poor performers, as determined by the median split. **B.** A series of scatter plots showing Pearson correlation coefficients among internal SNR, early SMG, late IFG activation, and behavioral accuracy. **C.** A scatter plot showing the regression coefficients from a linear regression model where behavioral accuracy is the dependent variable while internal SNR, early SMG, late IFG activation are the predictor variables. *The linear model significantly predicts behavioral accuracy while internal SNR is the only significant predictor.

To address our primary research question, we evaluated the simultaneous contribution of primary auditory pathways and frontal compensatory processes. For this, we computed internal SNR as the amplitude ratio of the noise and target related cortical evoked responses (e.g., **Figure 6A**), expecting to quantify an individual's speech unmasking ability. We also computed the mean activity in SMG and IFG separately averaged over 150 to 350 ms (the early component), and 500 to 800 ms (late).

Figure 6B shows correlations among internal SNR, SMG/IFG source activation, and SiN performance (accuracy) in the low SNR condition, that are obtained from 26 leave-one-out grand averages using jackknifing approach (Luck, 2014; Stahl & Gibbons, 2004). Greater

internal SNR predicted better performance (R = 0.45, p = 0.022). Stronger early SMG activation also predicted better performance (R = 0.39, p = 0.047). However, greater late IFG activity predicted poorer performance (R = -0.42, p = 0.033). This suggests that both internal SNR and SMG activation are positively related to SiN performance, while IFG activation has negative relation to the performance.

To test their joint contributions, we conducted a linear regression in which internal SNR, SMG, IFG activation were simultaneously related to SiN performance (**Figure 6C**). As a whole, these factors accounted for a large proportion of the variance (R = 0.65, p = 0.00033). However, among these three factors, internal SNR was the only significant predictor (t = 2.25, df = 22, p = 0.035), whereas SMG (t = 1.80, df = 22, p > 0.05) and IFG activation (t = -2.03, df = 22, p > 0.05) did not significantly contribute in the prediction SiN performance. This suggests that internal SNR (representing the contribution of lower level signal analysis) uniquely predicts variation in accuracy.

Discussion

We found SNR effects on timing and location of cortical activity for a speech-in-noise recognition task. In a relatively easy SNR condition (in which subjects achieved ~80% accuracy), left SMG showed a relatively early evoked response (~250 ms after target word onset). In contrast, a challenging SNR (~68% accuracy) elicited the late response in left IFG (~700 ms after target word onset). Within the same “external” SNR condition, individual differences in such SMG and IFG responses predicted SiN performance; good performers showed stronger early SMG response while poor performers showed stronger late IFG activity. Individual differences in the ratio of noise- to target word-evoked cortical responses—the “internal SNR,”—also predicted SiN performance; subjects with lower internal SNR exhibited poorer accuracy. Importantly, both SMG and IFG responses did not contribute to the prediction of SiN performance when internal SNR was added to the linear regression model. These results

from correlational analyses could be explained by auditory scene analysis mechanisms for target unmasking and by temporal cortical processes for speech perception. A poorer ability to unmasking target speech from background noise may lead to increased frontal lobe processing that is employed when lower-level auditory pathways are unable to secure favorable speech quality due to background noise.

Internal SNR: A measure of pre-speech processing for speech unmasking

Pre-speech time cortical activity for speech unmasking should be localized to the primary and secondary AC, appearing as enhanced neural representation of the target sound (the speech) and suppressed neural representation of ignored stimuli (the noise). This is consistent with work suggesting that the auditory N1 (the numerator of the measure) can be localized to AC and the planum temporale (Schneider et al., 2002).

Such responses could reflect auditory selective attention, which shows a similar pattern in previous studies (Hillyard et al., 1973; Hillyard et al., 1998; Mesgarani and Chang, 2012). In the present study, good performers showed significantly weaker noise-evoked AC response, compared with poor performers, approximately 200 ms after the noise onset (**Figure 6A**). Decreased AC response to background noise in good performers is compatible with the presence of a sensory gain control mechanism (Hillyard et al., 1998). The variation in the sensory gain control may originate from multiple factors. It may reflect the acuity of encoding spectro-temporal acoustic cues from speech and noise or grouping of such acoustic cues for auditory object formation. It may also reflect endogenous mechanisms for active suppression of background sounds along with neural enhancement of foreground sounds (Shinn-Cunningham and Best, 2008).

Our goal was not to disentangle the sources of variation in sensory gain control but rather to quantify the effectiveness of sensory gain control by internal SNR and test how it predicts later speech processes and behavioral accuracy. In this regard, we found a significant

correlation between accuracy and the relative amplitude word- and noise-evoked potentials. This demonstrates that individuals have a differential ability to suppress the noise effectively via early auditory processes (indexed by internal SNR). The first-order correlations between IFG and SMG activation and behavioral performance were compatible with the effect of SNR on ERPs: Good performers had stronger SMG activation (as in the high SNR condition), while poor performers had stronger IFG activation (as the low SNR condition). However, the amplitude of SMG and IFG response did not uniquely contribute to accuracy when internal SNR was added to the model. This indicates that changes in SMG or IFG activity are the outcome of pre-speech sensory gain control processing, rather than an independent causal factor predicting speech perception performance.

This result conflicts with some findings from earlier studies but also clarifies their findings. For example, Wong et al. (2009) did not find a relationship between SiN performance and AC or auditory related cortex. However, this study used fMRI, which may have missed the contribution of much shorter-lived bottom-up processes (Parbery-Clark et al., 2009). Like us, they did find correlations with activity in the IFG (and also the precentral gyrus a second dorsal route site). However, this study included both younger and older adults (who showed differences in cortical networks). Thus, some of these correlations may have been driven by age differences. Also, it is unclear whether these differences in cortical activity are necessary for successful SiN understanding (or at least helpful), as such differences could also reflect processes like increased effort (consistent with the view that IFG may be a domain-general control process) (Fedorenko et al., 2013), error monitoring, or even just increased uncertainty. Such processes may be engaged by noise without necessarily playing a causal role in improving perception.

Similarly, Bidelman and Howell (2016) related both AC and dorsal route activity to performance. They found no contribution of AC, but a correlation between speech performance and early (~115 ms) activity in the IFG. However, their measure of AC activity represented the

response to both the speech and noise, not the ability of AC to pull speech from the noise. Moreover, speech perception accuracy was assessed in an offline task. As a result, the cortical activity measures did not reflect cortical processes leading up to accurate (or inaccurate response); thus, these correlations may reflect broader individual differences, rather than a causal chain leading to accurate SiN processing.

SNR effect on timing and location of cortical activity

Previous studies have suggested that spoken-word recognition occurs via a process of dynamic lexical competition as speech unfolds over time. For many words, this competition begins to resolve (e.g., the target separates from competitors) around 300 ms after word onset (Huettig and Altmann, 2005; Farris-Trimble and McMurray, 2013). In significantly challenging conditions (high noise) however, lexical processing can be delayed about 250 ms until most of the word has been heard (Farris-Trimble et al., 2014; McMurray et al., 2017), which may minimize competition. Based on the timing predicted by these studies, we expected early SMG processing in the high SNR condition, and late IFG processing in the low SNR condition.

Our secondary analysis, a data-driven approach based on spatiotemporal clustering analysis of ERPs (microstate analysis), supports the conclusion from the ROI-based analysis. As microstates 1 and 2 explained the greatest amount of the signal's variance in the low and the high SNR condition, respectively, we focused on the highest GFP peak timepoints, within corresponding microstates for each SNR condition. Whole brain maps obtained from those timepoints were supportive of the ROI analysis: in the high SNR condition, SMG was strongly activated in the left hemisphere, while left IFG and bilateral Heschl's gyrus (HG) were activated in the low SNR condition.

Increased SMG activity between the second and the third phonemes (see **Figure 4**) in the high SNR condition may indicate a neural substrate of immediate lexical access (Farris-Trimble et al., 2014; McMurray et al., 2017), consistent with Gow (2012). This immediacy was

observed when speech sounds were relatively clean (high SNR), and it does not appear in previous EEG studies using non-word synthesized phonemes (Bidelman and Dexter, 2015; Bidelman and Howell, 2016). However, we note that Bidelman and Howell used a single non-word stimulus (a vowel-consonant-vowel) that would not be expected to engage lexical processing. Bidelman and Howell's results also demonstrated an early activity (~115 ms) in IFG with a clearly intelligible VCV phoneme. This was not observed in our study. However, because we used naturally spoken CVC words, we can limit the interpretation of the late IFG activity to the decision-making process in which listeners are trying to clean up the results of lexical competition in SMG.

The idea that greater IFG activity is linked with poorer SiN recognition performance seems to be inconsistent with some fMRI studies that showed the positive correlation between the IFG activity and speech recognition performance (Zekveld et al., 2006; Wong et al., 2009; Vaden et al., 2015; Du et al., 2016). This may stem from the difference between fMRI and EEG in temporal resolution and in sensitivity to either neural metabolic activity or the equivalent current dipoles (Bridwell et al., 2013). In the present study, we exhibited event-related potentials at a specific latency of ~700 ms after target word onset, i.e., ~200 ms after target word "offset." Previous fMRI studies might have demonstrated IFG activity at different latencies or accumulated BOLD signal that is not time-locked to a specific sensory event. Alternatively, this difference between fMRI and the current EEG results might be due to an error in the estimation of our source location. However, the fact that both of our approaches – an ROI-based approach and a data-driven approach without a space constraint – resulted in exhibiting the same IFG activity with a similar latency might make our results more reliable.

So, what is IFG contributing to the process? It is unknown whether these processes are due to active compensation for the noise or increased effort (both of which may help) or are a simply marker of increased response uncertainty. Our correlational analyses do not suggest a causal role for frontal activity in predicting an individual's accuracy. The first order correlation

was negative – more frontal activity was linked to *less* overall accuracy. More importantly, this correlation was not significant when internal SNR was added to the model suggesting IFG activity does not offer a unique contribution to accuracy. Rather it may simply reflect the clarity of the signal offered by the earlier auditory processes that deal with noise. That is, if we view IFG as primarily serving a decision-making role in this task when dorsal route areas do not output clear representations of the signal, IFG must work harder to resolve on a decision. It may be then that activity in frontal areas is not causally necessary for good SiN performance, but rather reflects the additional response uncertainty created by noisy listening situations. This challenges accounts like Du et al. (2016) that argue for a causal role of dorsal route processing in SiN understanding.

Methodological advances and justifications for source time course analysis

Our approach to identifying a single voxel within an ROI deserves a particular discussion. Identification of the representative voxel of an ROI is a problem common to EEG source analysis, fMRI, and other functional brain imaging studies. Many relevant neuroimaging analysis approaches have been described, including univariate, multivariate, and machine learning; however, most of these are intended for the identification of regions of interest or functional connections from a whole brain map. Drawbacks of this type of whole-brain analysis include the need for strict multiple comparisons correction and, therefore, decreased statistical power. Using strong a priori hypotheses to generate regions of interest allowed us to circumvent these issues, but still requires identification of representative voxels within our regions of interest. Favored approaches generally require identification of peak activity within an ROI (Tong et al., 2016). However, to avoid the assumption that choosing peak activity implies, we opted instead to choose the voxel that has the maximum average correlation to every other voxel within the ROI. In the present study, we chose not to constrain the location of the voxel of interest within an ROI for each condition. Because our anatomic resolution is unlikely to be at

562 the voxel level, we elected to choose a different representative voxel for each condition,
563 unconstrained by the location of the representative voxel from other conditions.
564
565 **Conclusion**
566 We found that clean, intelligible speech elicits early processing at SMG, while sensory
567 degradation results in late processing at IFG for less intelligible speech. Better speech
568 unmasking in good performers modulated the ratio of cortical evoked responses to the
569 background noise and target sound, which effectively changed SNR internally, resulting in
570 facilitated lexical/phonological processing through SMG. These findings may collectively form a
571 neural substrate of individual differences in speech-in-noise understanding ability. Crucially,
572 however, only neural representation of SNR uniquely predicted variation in performance,
573 suggesting that individual differences in SiN comprehension are largely a matter of primary
574 processes that extract the signal from noise rather than later compensatory ones.
575

576 **References**

577 Allopenna PD, Magnuson JS, Tanenhaus MK (1998) Tracking the Time Course of Spoken Word
578 Recognition Using Eye Movements: Evidence for Continuous Mapping Models. Journal
579 of memory and language 38:419-439.
580 Ben-David BM, Chambers CG, Daneman M, Pichora-Fuller MK, Reingold EM, Schneider BA
581 (2011) Effects of aging and noise on real-time spoken word recognition: evidence from
582 eye movements. Journal of speech, language, and hearing research : JSLHR 54:243-
583 262.
584 Bidelman GM, Dexter L (2015) Bilinguals at the "cocktail party": dissociable neural activity in
585 auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners'
586 speech-in-noise recognition deficits. Brain Lang 143:32-41.
587 Bidelman GM, Howell M (2016) Functional changes in inter- and intra-hemispheric cortical
588 processing underlying degraded speech perception. Neuroimage 124:581-590.
589 Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory
590 and decision processes in auditory object identification. Nat Neurosci 7:295-301.
591 Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433-436.
592 Bregman AS (1999) Auditory scene analysis : the perceptual organization of sound. Cambridge,
593 Mass.: MIT Press.
594 Bridwell DA, Wu L, Eichele T, Calhoun VD (2013) The spatiospectral characterization of brain
595 networks: fusing concurrent EEG spectra and fMRI maps. Neuroimage 69:101-111.
596 Brouwer S, Bradlow AR (2016) The Temporal Dynamics of Spoken Word Recognition in
597 Adverse Listening Conditions. Journal of psycholinguistic research 45:1151-1160.
598 Dahan D, Gareth Gaskell M (2007) The temporal dynamics of ambiguity resolution: Evidence
599 from spoken-word recognition. Journal of memory and language 57:483-501.
600 Dale AM, Liu AK, Fischl BR, Buckner RL, Belliveau JW, Lewine JD, Halgren E (2000) Dynamic
601 statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of
602 cortical activity. Neuron 26:55-67.
603 Du Y, Buchsbaum BR, Grady CL, Alain C (2014) Noise differentially impacts phoneme
604 representations in the auditory and speech motor systems. Proc Natl Acad Sci U S A
605 111:7126-7131.
606 Du Y, Buchsbaum BR, Grady CL, Alain C (2016) Increased activity in frontal motor cortex
607 compensates impaired speech perception in older adults. Nat Commun 7:12241.
608 Farris-Trimble A, McMurray B (2013) Test-retest reliability of eye tracking in the visual world
609 paradigm for the study of real-time spoken word recognition. Journal of Speech
610 Language and Hearing Research 56:1328-1345.
611 Farris-Trimble A, McMurray B, Cigrand N, Tomblin JB (2014) The process of spoken word
612 recognition in the face of signal degradation. Journal of Experimental Psychology:
613 Human Perception and Performance Journal of Experimental Psychology: Human
614 Perception and Performance 40:308-327.
615 Fedorenko E, Duncan J, Kanwisher N (2013) Broad domain generality in focal regions of frontal
616 and parietal cortex. Proceedings of the National Academy of Sciences Proceedings of
617 the National Academy of Sciences 110:16616-16621.
618 Frey JN, Mainy N, Lachaux JP, Muller N, Bertrand O, Weisz N (2014) Selective Modulation of
619 Auditory Cortical Alpha Activity in an Audiovisual Spatial Attention Task. JOURNAL OF
620 NEUROSCIENCE 34:6634.
621 Friston K, Harrison L, Daunizeau J, Kiebel S, Phillips C, Trujillo-Barreto N, Henson R, Flandin
622 G, Mattout J (2008) Multiple sparse priors for the M/EEG inverse problem. Neuroimage
623 39:1104-1120.
624 Gagne JP, Besser J, Lemke U (2017) Behavioral Assessment of Listening Effort Using a Dual-
625 Task Paradigm. Trends Hear 21:1-25.

626 Goldberg HR, Choi I, Varghese LA, Bharadwaj H, Shinn-Cunningham BG (2014) Auditory
627 attention in a dynamic scene: Behavioral and electrophysiological correlates. The
628 Journal of the Acoustical Society of America The Journal of the Acoustical Society of
629 America 135:2415.
630 Gow DW, Jr. (2012) The cortical organization of lexical knowledge: a dual lexicon model of
631 spoken language processing. Brain Lang 121:273-288.
632 Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L,
633 Hamalainen MS (2014) MNE software for processing MEG and EEG data. Neuroimage
634 86:446-460.
635 Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M,
636 Brooks T, Parkkonen L, Hamalainen M (2013) MEG and EEG data analysis with MNE-
637 Python. Front Neurosci 7:267.
638 Hämäläinen MS, Sarvas, J. (1989) Realistic conductivity geometry model of the human head for
639 interpretation of neuromagnetic data. IEEE Trans Biomed Eng 36:165-171.
640 Harris RW, Swenson DW (1990) Effects of reverberation and noise on speech recognition by
641 adults with various amounts of sensorineural hearing impairment. Audiology : official
642 organ of the International Society of Audiology 29:314-321.
643 Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci
644 8:393-402.
645 Hillyard SA, Vogel EK, Luck SJ (1998) Sensory gain control (amplification) as a mechanism of
646 selective attention: electrophysiological and neuroimaging evidence. Philosophical
647 transactions of the Royal Society of London Series B, Biological sciences 353:1257-
648 1270.
649 Hillyard SA, Hink RF, Schwent VL, Picton TW (1973) Electrical signs of selective attention in the
650 human brain. Science 182:177-180.
651 Huettig F, Altmann GT (2005) Word meaning and the control of eye fixation: semantic
652 competitor effects and the visual world paradigm. Cognition 96:B23-32.
653 Kisler T, Reichel UD, Sciel F (2017) Multilingual processing of speech via web services.
654 Computer Speech & Language 45:326-347.
655 Koenig T, Kottlow M, Stein M, Melie-Garcia L (2011) Ragu: a free tool for the analysis of EEG
656 and MEG event-related scalp field data using global randomization statistics. Comput
657 Intell Neurosci 2011:1-14.
658 Koenig T, Stein M, Grieder M, Kottlow M (2014) A tutorial on data-driven methods for
659 statistically assessing ERP topographies. Brain Topogr 27:72-83.
660 Lange K (2009) Brain correlates of early auditory processing are attenuated by expectations for
661 time and pitch. Brain and cognition 69:127.
662 Lee AKC, Rajaram S, Xia J, Bharadwaj H, Larson E, Hämäläinen MS, Shinn-Cunningham BG
663 (2013) Auditory Selective Attention Reveals Preparatory Activity in Different Cortical
664 Regions for Selection Based on Source Location and Source Pitch. Front Neurosci
665 Frontiers in Neuroscience 6.
666 Lehmann D (1989a) Microstates of the Brain in EEG and ERP Mapping Studies.72-83.
667 Lehmann D (1989b) From Mapping to the Analysis and Interpretation of EEG/EP Maps.53-75.
668 Liberman MC, Epstein MJ, Cleveland SS, Wang H, Maison SF (2016) Toward a differential
669 diagnosis of hidden hearing loss in humans. PLoS One 11:1-16.
670 Liebenthal E, Sabri M, Beardsley SA, Mangalathu-Arumana J, Desai A (2013) Neural dynamics
671 of phonological processing in the dorsal auditory stream. J Neurosci 33:15414-15424.
672 Luck SJ (2014) An introduction to the event-related potential technique, Second edition. Edition.
673 Cambridge, Massachusetts: The MIT Press.
674 Magnuson JS, Dixon JA, Tanenhaus MK, Aslin RN (2007) The dynamics of lexical competition
675 during spoken word recognition. Cognitive science 31:133-156.

McMurray B, Farris-Trimble A, Rigler H (2017) Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally. Cognition 169:147-164.

McQueen JM, Huettig F (2012) Changing only the probability that spoken words will be distorted changes how they are recognized. The Journal of the Acoustical Society of America The Journal of the Acoustical Society of America 131:509-517.

Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:233-236.

Myers EB, Blumstein SE, Walsh E, Eliassen J (2009) Inferior Frontal Regions Underlie the Perception of Phonetic Category Invariance. PSCI Psychological Science 20:895-903.

Nabelek AK (1988) Identification of vowels in quiet, noise, and reverberation: relationships with age and hearing loss. The Journal of the Acoustical Society of America 84:476-484.

Ohlenforst B, Zekveld AA, Lunner T, Wendt D, Naylor G, Wang Y, Versfeld NJ, Kramer SE (2017) Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. Hear Res 351:68-79.

Ott CG, Langer N, Oechslin MS, Meyer M, Jancke L (2011) Processing of voiced and unvoiced acoustic stimuli in musicians. Front Psychol 2:195.

Owens E, Schubert ED (1977) Development of the California Consonant Test. J Speech Lang Hear Res 20:463-474.

Parbery-Clark A, Skoe E, Kraus N (2009) Musical experience limits the degradative effects of background noise on the neural processing of sound. J Neurosci 29:14100-14107.

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437-442.

Schiller B, Gianotti LR, Baumgartner T, Nash K, Koenig T, Knoch D (2016) Clocking the social mind by identifying mental processes in the IAT with electrical neuroimaging. Proc Natl Acad Sci U S A 113:2786-2791.

Schneider P, Scherg M, Dosch HG, Specht HJ, Gutschalk A, Rupp A (2002) Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. Nature neuroscience 5:688-694.

Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. Trends Neurosci 26:100-107.

Shinn-Cunningham BG, Best V (2008) Selective attention in normal and impaired hearing. Trends Amplif 12:283-299.

Stahl J, Gibbons H (2004) The application of jackknife-based onset detection of lateralized readiness potential in correlative approaches. Psychophysiology 41:845-860.

Storey JD (2002) A direct approach to false discovery rates. JOURNAL- ROYAL STATISTICAL SOCIETY SERIES B STATISTICAL METHODOLOGY 64:479-498.

Tong Y, Chen Q, Nichols TE, Rasetti R, Callicott JH, Berman KF, Weinberger DR, Mattay VS (2016) Seeking optimal region-of-interest (ROI) single-value summary measures for fMRI studies in imaging genetics. PLoS One 11:1-20.

Vaden KI, Jr., Kuchinsky SE, Ahlstrom JB, Dubno JR, Eckert MA (2015) Cortical activity predicts which older adults recognize speech in noise and when. J Neurosci 35:3929-3937.

Wackermann J, Lehmann D, Michel CM, Strik WK (1993) Adaptive segmentation of spontaneous EEG map series into spatially defined microstates. INTPSY</cja:jid> International Journal of Psychophysiology 14:269-283.

Wong PC, Uppunda AK, Parrish TB, Dhar S (2008) Cortical mechanisms of speech perception in noise. Journal of speech, language, and hearing research : JSLHR 51:1026-1041.

Wong PC, Jin JX, Gunasekera GM, Abel R, Lee ER, Dhar S (2009) Aging and cortical mechanisms of speech perception in noise. Neuropsychologia 47:693-703.

726 Zekveld AA, Heslenfeld DJ, Festen JM, Schoonhoven R (2006) Top-down and bottom-up
727 processes in speech comprehension. Neuroimage 32:1826-1836.
728

729

730 **Figure Legends**

731 **Figure 1.** Trial and stimulus structure. Every trial starts with the cue phrase "check the word." A
732 target word starts 1 second after the noise onset. Four choices are given after the word ends;
733 subjects select the correct answer with a keypad. No feedback is given. The noise level is
734 manipulated to create high (+3 dB) and low (-3 dB) SNR conditions. Subjects complete 50 trials
735 for each condition.
736

737 **Figure 2.** Behavioral results. **A**. Summary of behavioral performance for the two conditions (+3
738 and -3 dB SNR). Boxes denote the 25th – 75th percentile range; the horizontal bars in the
739 center denote the median; the ranges are indicated by vertical dashed lines. Solid lines connect
740 points for the same subject in different conditions. **B**. Average accuracy as a function of reaction
741 time in +3 dB SNR condition. **C**. Average accuracy and reaction time in -3 dB SNR condition.
742

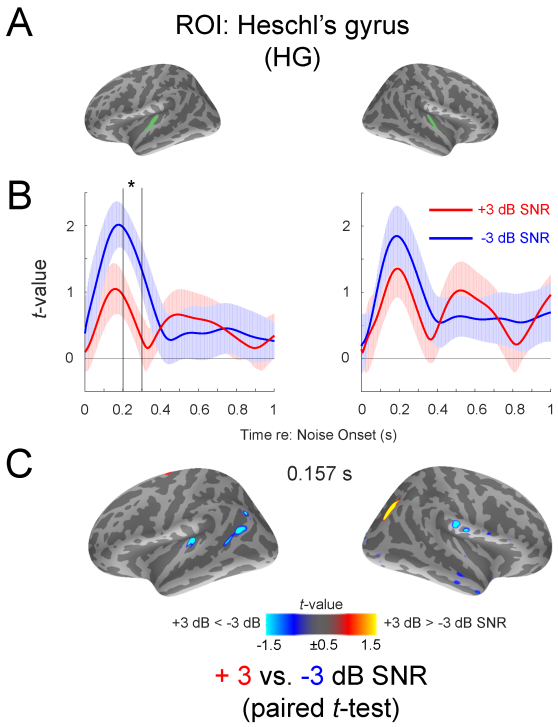
743 **Figure 3**. Region-of-interest (ROI) based source analysis. **A**. Cortical labels for two ROIs in left
744 and right hemispheres: supramarginal gyrus (SMG), and the pars opercularis and triangularis of
745 the inferior frontal gyrus (IFG), respectively. **B**. The time course of the t-value envelope, with the
746 standard error of the mean (±1 SEM), obtained at representative voxels in each SNR condition
747 (red color: +3 dB SNR, blue color: -3 dB SNR). An asterisk shows the timing of significant
748 difference between +3 and -3 dB SNR conditions (paired t-test, FDR adjusted $p < 0.05$). **C**.
749 Whole brain maps showing statistical contrasts (t-values obtained from paired t-tests between
750 the two SNR conditions) of source activation at each voxel, only displaying those with p-value
751 less than 0.05, at the timepoint that shows significant differences over the broadest area in the
752 ROIs within the time range described above.
753

754 **Figure 4.** Timings of significant cortical activity relative to distributions of phonological events.
755 **A**. Top and second panel show a histogram of the onsets of second and final phoneme of each
756 stimulus. The third panel shows superimposed temporal envelopes extracted from waveforms of
757 the 100 words. **B**. The whole brain maps at the bottom are from **Figure 3C** that shows statistical
758 contrasts of source activation at the timepoints that show significant differences between the
759 two SNR conditions. Purple curves on the cortical maps represent the conceptual illustration of
760 ascending information flow through the dorsal pathway.
761

762 **Figure 5**. Microstate analysis. **A**. Evoked responses over time after the word onset for +3 and -3
763 dB SNR condition. Each color represents ERPs from a different channel of interest. **B**. Global
764 field power (GFP) is calculated at each timepoint that is assigned to one of the microstate
765 clusters. **C**. Four microstate cluster maps. Dark blue, light blue, red, and dark pink colors
766 represent microstates 1, 2, 3, and 4, respectively. The relative area of GFP is calculated and
767 reveals the highest value for the microstate 1 and 2 for -3 and +3 dB SNR condition,
768 respectively. **D**. Whole brain maps obtained at the times assigned to microstates 1 and 2, that
769 show maximum GFP and the maximum peak of ERPs at the frontal-central electrodes.
770

771 **Figure 6.** Individual differences in speech-in-noise processing. **A**. Global field power of the
772 grand mean evoked potentials after the noise onset and after the target word onset, separately
773 in the low SNR condition. Scalp topographies were examined at the timepoints, suggested by
774 microstate analysis from **Figure 5**, and compared between good and poor performers, as

775 determined by the median split. **B**. A series of scatter plots showing Pearson correlation
776 coefficients among internal SNR, early SMG, late IFG activation, and behavioral accuracy. **C**. A
777 scatter plot showing the regression coefficients from a linear regression model where behavioral
778 accuracy is the dependent variable while internal SNR, early SMG, late IFG activation are the
779 predictor variables. *The linear model significantly predicts behavioral accuracy while internal
780 SNR is the only significant predictor.
781



Supplement Figure 1. Region-of-interest (ROI) based source analysis. **A.** Cortical labels for Heschl's gyrus in left and right hemispheres. **B.** The time course of the *t*-value envelope, with the standard error of the mean (±1 SEM), obtained at representative voxels in each SNR condition (red color: +3 dB SNR, blue color: -3 dB SNR). An asterisk shows the timing of significant difference between +3 and -3 dB SNR conditions (paired *t*-test, FDR adjusted $p < 0.05$). **C.** Whole brain maps showing statistical contrasts (*t*-values obtained from paired *t*-tests between the two SNR conditions) of source activation at each voxel, only displaying those with *p*-value less than 0.05.

782